ФЕДЕРАЛЬНАЯ ЦЕЛЕВАЯ ПРОГРАММА
"ГОСУДАРСТВЕННАЯ ПОДДЕРЖКА ИНТЕГРАЦИИ ВЫСШЕГО ОБРАЗОВАНИЯ
И ФУНДАМЕНТАЛЬНОЙ НАУКИ"

НОВЫЕ РАЗДЕЛЫ ФИЗИКИ ПОЛУПРОВОДНИКОВ

**А. Я. Шик, Л. Г. Бакуева,
С. Ф. Мусихин, С. А. Рыков**

ФИЗИКА НИЗКОРАЗМЕРНЫХ СИСТЕМ

Под общей редакцией
В. И. Ильина и А. Я. Шика

*Рекомендовано Министерством образования Российской Федерации
в качестве учебного пособия для студентов высших
учебных заведений, обучающихся по направлению
«Техническая физика»*



Санкт-Петербург
«Наука»
2001

УДК 621.38
ББК 22.379
Ш 76

Издание осуществлено при финансовой поддержке Федеральной целевой программы «Государственная поддержка интеграции высшего образования и фундаментальной науки»

Шик А. Я., Бакуева Л. Г., Мусихин С. Ф., Рыков С. А. Физика низкоразмерных систем / Под ред. А. Я. Шика. — СПб.: Наука, 2001. 160 с. 69 ил. (Серия учебных пособий «Новые разделы физики полупроводников»).

ISBN 5-02-024966-1

В пособии рассмотрены основные физические свойства систем с пониженной размерностью: полупроводниковых структур с двумерным электронным газом, квантовых нитей и квантовых точек. Изложены принцип размерного квантования и условия наблюдения квантово-размерных явлений, рассмотрены особенности функции плотности состояний и статистики носителей заряда в низкоразмерных системах, оптические свойства квантовых ям, кинетические эффекты, в том числе в магнитных полях, физические явления, наблюдаемые только в таких системах. Описаны основные технологические методы получения квантово-размерных структур и даны примеры практического использования таких структур в наноэлектронике.

Учебное пособие предиазиачено для студентов старших курсов, изучающих физику полупроводников и приборы наноэлектроники; оно также может быть рекомендовано аспираитам и научным работникам, занимающимся фундаментальными и прикладными исследованиями в области физики и техники низкоразмерных структур.

Рецензенты:

д-р физ.-мат. наук *В. А. МОШНИКОВ*,
д-р физ.-мат. иаук *Т. А. ПОЛЯНСКАЯ*

ISBN 5-02-024966-1

ВВЕДЕНИЕ

Последние три десятилетия развития физики твердого тела характеризуются тем, что основными объектами исследования все в большей степени становятся не массивные кристаллы, а тонкие пленки, многослойные тонкопленочные системы, проводящие нити и кристаллиты малого размера. В таких системах существенно меняется большинство электронных свойств — возникает большое число новых, так называемых размерных эффектов. Наиболее кардинальной перестройкой свойств отличаются квантовые размерные структуры, в которых свободные носители заряда локализованы в одном, двух или во всех трех координатных направлениях в области с размерами порядка дебройлевской длины волны носителей. При этом вступают в силу законы квантовой механики и происходит изменение наиболее фундаментальной характеристики электронной системы — ее энергетического спектра. Спектр становится дискретным для движения вдоль координаты, по которой ограничено движение. Если движение ограничено вдоль одного или двух направлений, то под влиянием внешних полей и взаимодействий с рассеивателями (фононы, примеси) могут меняться уже не три, а лишь две или только одна из компонент импульса электронов и дырок, в результате чего носители ведут себя как двумерный или одномерный газ. Квантовые структуры, в которых движение носителей ограничено во всех трех направлениях, напоминают искусственные атомы. Здесь энергетический спектр является чисто дискретным.

Квантово-размерные структуры обладают целой совокупностью уникальных свойств, весьма далеких от того, что можно наблюдать в системе обычных, трехмерных электронов и дырок. Наиболее ярким проявлением этого является квантовый эффект Холла в двумерных системах, открытый в 1980 г. и всего пять лет спустя принесший своему первооткрывателю Нобелевскую премию. Такие структуры могут служить основой для создания новых типов полупроводниковых приборов, в первую очередь для опто- и наноэлектроники.

То обстоятельство, что квантово-размерные структуры находятся в центре внимания именно сейчас, вызвано интенсивным развитием в последние годы технологии изготовления полупроводниковых структур — молекулярно-лучевой эпитаксии, газофазной эпитаксии, нанолитографии, открытием явления самоорганизации наноструктур. Это дает возможность создания такого рода структур любого профиля с точностью до одного атомного слоя.

К настоящему времени физика низкоразмерных электронных систем, с одной стороны, продолжает интенсивно развиваться и находиться на самом передовом рубеже современной физики твердого тела, а с другой стороны, эта дисциплина во многих своих разделах уже сформировала некоторую сложившуюся систему представлений, что позволяет с уверенностью вводить их в вузовский курс без опасения, что завтра они будут опровергнуты стремительным развитием науки.

Необходимость в данном пособии продиктована тем, что на сегодняшний день весь огромный массив литературы по квантово-размерным системам представляет собой разрозненный набор статей в различных журналах. Существует лишь несколько отдельных обзоров по некоторым частным проблемам, которые не охватывают все то, что следовало бы отразить в курсах лекций, а, кроме того, зачастую эти сведения опубликованы в литературе, недоступной массовому российскому читателю. Монографии и тем более учебники по указанной теме на русском языке отсутствуют. Краткое популярное введение в физику низкоразмерных систем дано в статьях энциклопедии «Современное естествознание».[1]

Главы 1, 3, 4 и 6 написаны А. Я. Шиком, глава 2 — С. Ф. Мусихиным и Л. Г. Бакуевой, главы 5 и 7 — А. Я. Шиком и С. А. Рыковым, глава 8 — А. Я. Шиком, С. Ф. Мусихиным и С. А. Рыковым.

---

[1] *Демиховский В. Я.* Квантово-размерные эффекты // Современное естествознание: Энциклопедия. В 10-ти т. Т. 5. Физика конденсированных сред. М.: Изд. дом Магистр-пресс, 2000. С. 140—146; *Шик А. Я.* Двумерные электронные системы // Там же. С. 147—153; *Шик А. Я.* Квантовые нити // Там же. С. 154—159.

Глава 1

РАЗМЕРНОЕ КВАНТОВАНИЕ И КВАНТОВО-РАЗМЕРНЫЕ СТРУКТУРЫ

Весь комплекс явлений, обычно понимаемый под словами «электронные свойства низкоразмерных электронных систем» и описываемый в данной книге, многочисленные новые типы электронных приборов, использующих эти свойства (см. главу 8), — все это имеет в основе один фундаментальный физический факт: изменение энергетического спектра электронов и дырок в структурах с очень малыми размерами. В данной главе мы рассмотрим физику этого явления, называемого обычно размерным квантованием, условия его наблюдения и классификацию размерно-квантованных структур.

1.1. Принцип размерного квантования

Мы продемонстрируем основную идею размерного квантования на примере электронов, находящихся в очень тонкой металлической или полупроводниковой пленке толщиной a [1]. То обстоятельство, что в обычных условиях носители сосредоточены в пленке и не выходят из нее в окружающую среду, означает, что материал пленки (металл или полупроводник) представляет собой потенциальную яму для электронов глубиной, равной работе выхода Φ, и шириной a. Согласно законам квантовой механики, энергия электронов в такой яме квантуется, т. е. может принимать лишь некоторые дискретные значения E_n, где n имеет целочисленные значения 1, 2, 3,... (весьма часто в литературе нумерацию n начинают не с единицы, а с нуля). Эти дискретные значения называют уровнями размерного квантования.

Типичные значения работы выхода в большинстве твердых тел имеют величину Φ = 4—5 эВ, на несколько порядков превышающую характерную тепловую энергию носителей kT, равную при комнатной температуре 0.026 эВ. Поэтому в нашем примере потенциальную яму можно считать бесконечно глубокой (рис. 1.1, *а*). Если пленка занимает область $0 < z < a$, то в этом случае волновые

функции и энергетические уровни квантовых состояний даются простыми формулами:

$$\psi_n(z) = \sqrt{2/a}\sin\frac{\pi n}{a}z, \tag{1.1}$$

$$E_n = \frac{\pi^2\hbar^2 n^2}{2ma^2}, \tag{1.2}$$

где m — эффективная масса электрона. Другим необходимым условием, позволяющим считать яму бесконечно глубокой, является малость значений E_n, даваемых формулой (1.2), по сравнению с ее действительной глубиной Ф. Это условие, которое для нижних уровней можно записать в виде $a > \pi\hbar/(2m\Phi)^{1/2}$ при $m = 0.1m_0$, соответствует толщинам пленки более 1 нм, что соответствует нескольким межатомным расстояниям. Во всех реальных структурах это условие соблюдается.

Мы привели формулу (1.2) для примера тонкой пленки. Ниже в данной главе будут рассмотрены и другие типы квантово-размерных структур, где движение носителей также ограничено некоторой малой областью с характерными размерами a. При этом точная формула (1.2) уже неприменима, но сам вывод о наличии квантования, разумеется, остается в силе, поскольку вытекает из общих принципов квантовой механики. Сохраняет силу и порядковая оценка для энергий квантовых уровней

$$E_n \sim \frac{\hbar^2}{ma^2}, \tag{1.3}$$

являющаяся следствием гейзенберговского соотношения неопределенностей и включающая (1.2) в качестве частного случая.

Вывод о квантовании энергии электронного движения и приведенные формулы для E_n относятся лишь к движению поперек потенциальной ямы (по оси z). На движение в плоскости xy (параллельно границам пленки) потенциал ямы не влияет. В этой плоскости носители движутся как свободные и характеризуются, как и в массивном образце, непрерывным квадратичным по импульсу энергетическим спектром с эффективной массой m. Полная же энергия носителей в квантово-размерной пленке носит смешанный дискретно-непрерывный спектр, представляя собой сумму дискретных уровней, связанных с движением в направлении квантования, и непрерывной компоненты, описывающей движение в плоскости слоя:

$$E = E_n + \frac{p_x^2 + p_y^2}{2m}, \tag{1.4}$$

где p_x и p_y — компоненты импульса в плоскости слоя.

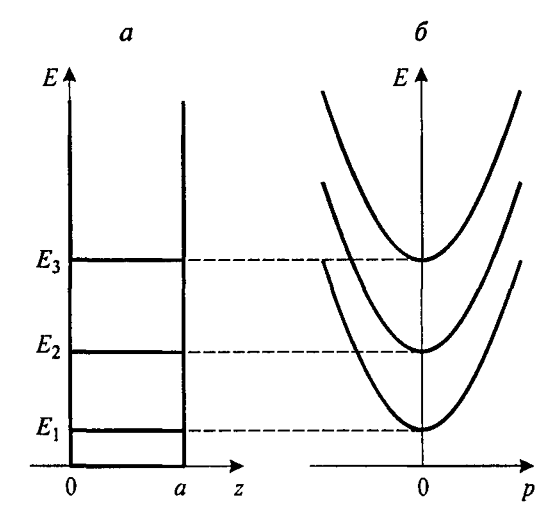


Рис. 1.1. Энергетический спектр квантово-размерной пленки ($p = \sqrt{p_x^2 + p_y^2}$ — импульс носителей заряда в плоскости пленки).

Заметим, что за счет непрерывной компоненты энергетического спектра $p^2/2m = (p_x^2 + p_y^2)/2m$ электроны, принадлежащие одному и тому же уровню E_n, могут иметь любую энергию от E_n до бесконечности (рис. 1.1, *б*). Такую совокупность состояний для данного фиксированного n обычно называют подзоной размерного квантования.

Пусть все или почти все электроны в системе имеют энергии, меньшие E_2, и потому принадлежат нижней подзоне размерного квантования (необходимые условия этого будут обсуждаться в главе 3). Тогда никакой упругий процесс (например, рассеяние на примесях или акустических фононах), равно как и рассеяние электронов друг на друге, не может изменить квантовое число n, переведя электрон на вышележащий уровень, поскольку это потребовало бы дополнительных затрат энергии. Это означает, что электроны при упругом рассеянии могут изменять только свой импульс в плоскости xy, т. е. ведут себя как чисто двумерные частицы. Поэтому квантово-размерные структуры, в которых заполнен лишь один квантовый уровень, часто называют двумерными электронными структурами.

Существуют и другие возможные квантовые структуры, где движение носителей ограничено не в одном, а в двух направлениях, как в микроскопической проволоке или нити (эти структуры обсуждаются в разделе 1.4). В этом случае носители могут свободно двигаться лишь в одном направлении, вдоль нити (назовем его осью x). В поперечном сечении (плоскость yz) энергия квантуется и принимает дискретные значения E_{mn} (как любое двумерное движение, оно описывается двумя квантовыми числами, m и n). Пол-

ный спектр при этом тоже является дискретно-непрерывным, но лишь с одной непрерывной степенью свободы:

$$E = E_{mn} + \frac{p_x^2}{2m}. \tag{1.5}$$

По аналогии с пленочными структурами, имеющими спектр вида (1.4.), данные системы называются одномерными электронными структурами или квантовыми нитями. Спектр квантовых нитей также представляет собой совокупность подзон размерного квантования, но уже не двумерных, а одномерных.

Наконец, существуют технологические возможности, обсуждаемые в разделе 1.5, а также в главе 2, создать квантовые структуры, напоминающие искусственные атомы, где движение носителей ограничено во всех трех направлениях [2]. Здесь энергетический спектр уже не содержит непрерывной компоненты, т. е. не состоит из подзон, а является чисто дискретным. Как и в атоме, он описывается тремя дискретными квантовыми числами (не считая спина) и может быть записан в виде $E = E_{lmn}$, причем, как и в атоме, энергетические уровни могут быть вырождены и зависеть лишь от одного или двух чисел. Подобные системы носят название нуль-мерных электронных структур или квантовых точек.

Из всего сказанного выше следует, что в твердотельных структурах, где хотя бы вдоль одного направления движение носителей ограничено очень малой областью, сравнимой по размерам с дебройлевской длиной волны носителей и характеризуемой обычно десятками нанометров, энергетический спектр носителей заметно меняется и становится частично или полностью дискретным. Подобное изменение спектра за счет размерного квантования приводит к существенному изменению всех электронных свойств системы по сравнению с массивным образцом того же материала.

1.2. Условия наблюдения квантовых размерных эффектов

Очевидно, что, для того чтобы описанное выше квантование энергетического спектра могло проявляться в каких-либо наблюдаемых эффектах, расстояние между энергетическими уровнями $E_{n+1} - E_n$ должно быть достаточно велико. В первую очередь оно должно значительно превосходить тепловую энергию носителей:

$$E_{n+1} - E_n \gg kT, \tag{1.6}$$

поскольку в противном случае практически одинаковая заселенность соседних уровней и частые переходы носителей между ними делают квантовые эффекты ненаблюдаемыми.

Если электронный газ вырожден и характеризуется энергией Ферми ζ, то желательно также выполнение условия

$$E_{n+1} - E_n \geq \zeta \tag{1.7}$$

(условие (1.6) при этом выполняется автоматически, поскольку для вырожденного газа $kT \ll \zeta$). При невыполнении указанного условия заполнено много квантовых уровней и квантовые размерные эффекты, будучи в принципе наблюдаемыми, имеют весьма малую относительную величину.

Существует еще одно необходимое требование для наблюдения квантовых размерных эффектов. В реальных структурах носители всегда испытывают рассеяние на примесях, фононах и др. Интенсивность рассеяния обычно характеризуется временем релаксации импульса τ, более подробно обсуждаемым в главе 5 и связанным прямой пропорциональностью с другой важной характеристикой носителей — их подвижностью $\mu = e\tau/m$. Величина τ представляет собой среднее время жизни в состоянии с данными фиксированными квантовыми числами (например, n, p_x, p_y для двумерного электронного газа). В силу соотношений неопределенности конечное значение τ влечет за собой неопределенность в энергии данного состояния $\Delta E \sim \hbar/\tau$. Очевидно, что говорить о наличии в системе отдельных дискретных уровней можно лишь в случае, когда расстояние между ними превышает неопределенность ΔE, т. е. при выполнении условия

$$E_{n+1} - E_n \gg \frac{\hbar}{\tau} = \frac{\hbar e}{m\mu}. \tag{1.8}$$

Можно показать, что выполнение условия (1.8) эквивалентно требованию того, чтобы длина свободного пробега носителей l значительно превосходила размер области a, в которой двигается носитель. Это достаточно очевидно. Согласно квантовой механике, квантование возникает при периодическом движении частицы. Это происходит лишь в случае достаточно слабого рассеяния, когда частица между двумя актами рассеяния (т. е. пройдя путь длиной l) успевает совершить несколько периодов колебаний, или, иными словами, несколько раз пересечь пленку (нить, точку) от границы до границы.

Поскольку расстояние между уровнями размерного квантования пропорционально $1/a^2$ (см. (1.3)), то из (1.6)—(1.8) следует, что для наблюдения квантовых размерных эффектов необходимы малые размеры структур, достаточно низкие температуры и высокие подвижности носителей, а также не слишком высокая их концентрация.

Приведем некоторые конкретные оценки. Чтобы наблюдать квантовые размерные эффекты в полупроводниках с $m = 0.1m_0$ (m_0 —

масса свободного электрона) при температурах вплоть до комнатной, необходимо иметь $a \leq 10$ нм. При этом подвижность носителей должна заметно превосходить величину 1000 см^2/(В·с). Если изготовить столь малые структуры не представляется возможным (проблемы технологии квантовых размерных структур обсуждаются в главе 2), то наблюдение квантовых эффектов возможно лишь при пониженных температурах и требует более высоких подвижностей носителей.

Заметим также, что в соответствии с требованием (1.7) металлические структуры мало подходят для наблюдения квантовых размерных эффектов, поскольку ζ в типичных металлах составляет несколько электронвольт, что заведомо больше любых расстояний между уровнями. Полупроводниковые или полуметаллические[1] структуры здесь явно предпочтительнее.

Еще одним важным условием, необходимым для наблюдения квантования, является высокое качество поверхностей, ограничивающих движение носителей в квантовых ямах, нитях и точках. Для тонких пленок речь идет о внешней границе пленки и о границе пленка—подложка. Для гетероструктур, обсуждаемых в последующих разделах, роль таких поверхностей играют гетеропереходы между различными полупроводниками. Характер отражения носителей от указанных границ должен быть близок к зеркальному, т. е. должен происходить с сохранением компоненты импульса, параллельной границе. Если это не так, то при каждом отражении от границы частица «забывает» о своем состоянии до отражения, т. е. на границе происходит эффективное рассеяние. Легко понять, что при этом длина пробега становится равной a и нарушается упомянутое выше условие $l \gg a$.

Для реализации зеркального отражения на границах необходимо, чтобы размеры шероховатостей, неизбежно существующих на любой поверхности, были меньше дебройлевской длины волны носителей. Кроме того, границы не должны содержать высокой плотности заряженных центров, приводящих к дополнительному рассеянию.

1.3. Структуры с двумерным электронным газом

Наиболее очевидным примером структур с двумерным электронным газом являются тонкие пленки, использованные в предыдущем разделе для иллюстрации идеи размерного квантования.

---

[1] Полуметаллами иазывают вещества, обладающие температурно-независимой (как металлы), ио достаточно низкой (10^{17}—10^{20} см$^{-3}$) концентрацией носителей. Наиболее распространенными полуметаллами явлются элементы V группы Bi, Sb, As.

Мы уже упоминали, что полупроводниковые или полуметаллические пленки являются для этой цели более предпочтительными, нежели металлические. И действительно, пленки Bi исторически были первым материалом, где экспериментально наблюдались квантовые размерные эффекты в проводимости [3] (об этом будет рассказано в главе 5). Такие пленки, имеющие необходимую толщину, достаточно высокую подвижность и хорошее качество поверхности, достаточно легко получаются методом вакуумного испарения.

Однако тонкие пленки не являются лучшим объектом для наблюдения квантовых эффектов. Упомянутые пленки Bi с их зеркально отражающими поверхностями — скорее исключение, обусловленное очень большой дебройлевской длиной волны в данном материале, чем правило. В других материалах, в том числе в полупроводниках, получить тонкие пленки необходимого качества весьма сложно. Причина в том, что на поверхности полупроводниковой пленки существует высокая плотность поверхностных состояний, играющих роль центров рассеяния. Поэтому тонкие пленки, широко использовавшиеся для изучения квантовых размерных эффектов в конце 60-х гг., уступили ведущую роль вначале кремниевым МДП-структурам, а впоследствии квантовым гетероструктурам.

Структуры типа МДП (металл—диэлектрик—полупроводник) были известны задолго до открытия квантовых размерных эффектов и использовались (и продолжают использоваться) в качестве полевых транзисторов как в дискретном, так и в интегральном исполнении. На рис. 1.2 изображена зонная диаграмма такой структуры. На металлический затворный электрод, отделенный слоем диэлектрика толщиной d, подается напряжение $V_з$, создающее в полупроводнике приповерхностный изгиб зон. Для достаточно больших $V_з$ этот изгиб может стать порядка ширины запрещенной зоны. При этом в полупроводнике вблизи границы с диэлектриком образуется тонкий инверсионный слой, содержащий носители противоположного знака, нежели в объеме полупроводника (в нашем случае слой n-типа в p-полупроводнике). Рассматривая металлический затвор и инверсионный слой как две обкладки плоского конденсатора, легко заключить, что двумерная плотность электронов в слое n_S (плотность состояний электронов на единицу площади двумерного электронного газа) будет пропорциональна напряжению на затворе:

$$n_S = \frac{\kappa_d}{4\pi e d}(V_з - V_0), \tag{1.9}$$

где κ_d — диэлектрическая проницаемость диэлектрика; V_0 — пороговое напряжение, соответствующее открытию инверсионного канала, т. е. появлению в нем электронов.

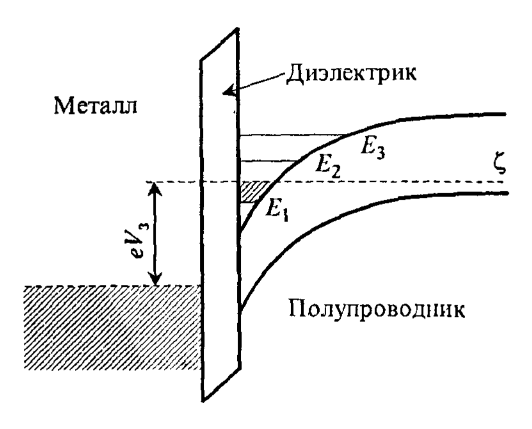


Рис. 1.2. Энергетическая диаграмма МДП-структуры.

Основным материалом для изготовления МДП-структур является кремний, в первую очередь благодаря той легкости, с которой путем окисления создается однородный слой высококачественного диэлектрика SiO_2, имеющий требуемую толщину.

Показанная на рис. 1.2 МДП-структура представляет собой готовый объект для наблюдения квантовых размерных эффектов [4]. Инверсионный слой представляет собой потенциальную яму для электронов, где одной стенкой является граница с диэлектриком, а роль второй стенки играет электростатический потенциал $e\Phi(z) = eFz$, прижимающий электроны к границе. Здесь $F \approx 4\pi e n_S/\kappa$ — электрическое поле в инверсионном слое, которое, согласно (1.9), пропорционально напряжению на затворе, а κ — диэлектрическая проницаемость полупроводника.

Легко оценить характерную толщину инверсионного канала Δz, играющую роль ширины квантовой ямы. Если определять Δz как ширину классически разрешенной области для электронов с энергией E_n, то $\Delta z = E_n/eF$. С другой стороны, согласно (1.3), $E_n \sim \hbar^2/m(\Delta z)^2$, что дает $\Delta z \sim F^{-1/3}$ и $E_n \sim F^{2/3}$.

Важнейшей особенностью МДП-структур, отличающей их от других квантово-размерных систем, является возможность управления концентрацией электронов n_S. Она может изменяться в широких пределах при изменении напряжения на затворе $V_з$ (см. (1.9)). Максимальное значение n_S определяется максимальным значением напряжения, которое можно приложить к затвору без риска пробоя диэлектрика. Для кремниевых структур оно имеет порядок 10^{13} см$^{-2}$.

Заметим, что изменение напряжения на затворе меняет одновременно концентрацию двумерных носителей n_S и расстояние между уровнями размерного квантования. Этим МДП-структура отличается, скажем, от тонкой пленки, где концентрация и энергия уров-

ней определяются соответственно уровнем легирования и толщиной пленки и могут меняться независимо.

Существует еще одно важное различие между МДП-структурами и тонкими пленками. Последние представляют собой потенциальную яму для обоих типов носителей, и квантование энергии (1.2) имеет место как для электронов, так и для дырок. В МДП-структуре, как и в любой иной системе, где ограничивающий потенциал (хотя бы с одной стороны) имеет электростатическую природу, квантуется энергия лишь одного типа носителей. Для другого типа носителей (на рис. 1.2 — дырок) потенциальная яма отсутствует и спектр остается непрерывным.

В настоящее время, однако, для изучения эффектов размерного квантования используются не МДП-, а гетероструктуры — контакты между полупроводниками с различной шириной запрещенной зоны. На таком контакте края энергетических зон испытывают скачки, ограничивающие движение носителей и играющие роль стенок квантовой ямы. На рис. 1.3 показана типичная зонная диаграмма гетероперехода между полупроводниками n- и p-типа (анизотипного перехода). Видно, что она сходна с диаграммой МДП-структуры. Так же как и там, в узкозонном полупроводнике вблизи границы раздела может образовываться инверсионный слой, играющий роль потенциальной ямы для электронов, в которой существуют уровни размерного квантования.

Важнейшим достоинством гетероперехода является высокое качество гетерограницы. При выборе в качестве компонент гетеропары вещества с хорошим согласием постоянных решетки (более подробно о соответствующих технологических проблемах рассказывается в главе 2) удается уменьшить плотность поверхностных состояний на гетерогранице до значений порядка 10^8 см$^{-2}$, что на несколько порядков меньше, чем в лучших МДП-структурах. Такая малая плотность состояний в совокупности с атомно-гладкой морфологией границы приводит к возможности получения рекордно высоких подвижностей μ в приповерхностном канале. В гетероструктурах GaAs–AlGaAs были получены значения подвижности электронов, превосходящие 10^7 см^2/(В · с), в то время как для лучших Si-МДП-структур $\mu \approx 5 \cdot 10^4$ см^2/(В · с). В результате столкновительное уширение уровней в гетероструктурах крайне мало, что позволяет наблюдать различные тонкие эффекты.

Концентрация носителей в канале гетероструктуры n_S определяется разрывами зон на гетерогранице и уровнями легирования компонент гетеропары. Для системы GaAs–AlGaAs она, как правило, не превосходит 10^{12} см$^{-2}$. В структуре, изображенной на рис. 1.3, эта величина является фиксированной. Если со стороны широкозонного материала на поверхность структуры нанести дополнительный за-

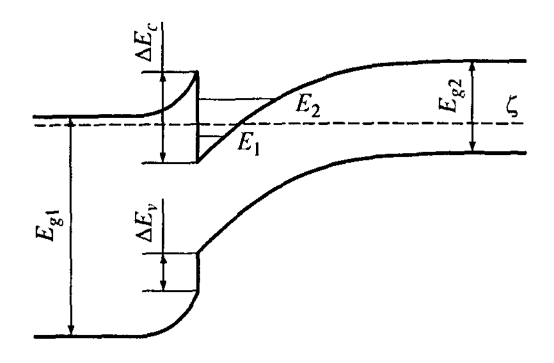


Рис. 1.3. Зонная диаграмма одиночного гетероперехода.

творный электрод, то, изменяя напряжение на нем, можно менять n_S в некоторых пределах, хотя и не столь эффективно, как в МДП-структурах.

В гетероструктуре рис. 1.3 потенциальная яма, ответственная за размерное квантование, образована разрывом зон ΔE_c с одной стороны и электростатическим полем перехода с другой. Это аналог инверсионного слоя МДП-структуры, в котором квантуется движение лишь одного типа носителей. Существуют также гетероструктурные аналоги тонкой пленки с размерным квантованием как электронов, так и дырок. Это двойные тонкослойные структуры, или квантовые ямы, представляющие очень тонкий слой узкозонного полупроводника между двумя широкозонными (рис. 1.4).

Если толщина узкозонного слоя a значительно меньше длины экранирования, то изгибы зон за счет объемного заряда в полупроводниках незначительны и зонная диаграмма структуры имеет вид, изображенный на рис. 1.4, *в*. Видно, что в зоне проводимости образуется практически прямоугольная квантовая яма шириной a и глубиной ΔE_c. Аналогичная яма глубиной ΔE_v существует в валентной зоне.

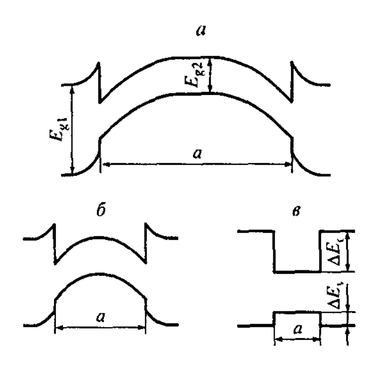


Рис. 1.4. Зонная диаграмма двойной гетероструктуры при различной толщине узкозонного слоя.

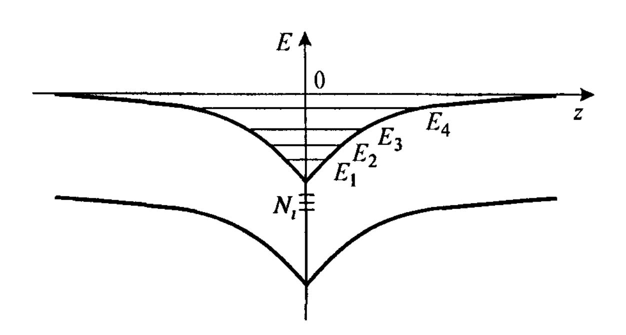


Рис. 1.5. Зонная диаграмма дельта-слоя.

Укажем в заключение еще на один тип квантово-размерных структур — так называемые дельта-слои [5]. Это полупроводники с предельно неоднородным профилем легирования, где примесные ионы не распределены однородно по объему, а сосредоточены в очень тонком слое в один или несколько периодов решетки. Носители, образовавшиеся при ионизации примесей, удерживаются их зарядом вблизи плоскости слоя. Электрическое поле слоя ионов экранируется зарядом электронов, и результирующая потенциальная яма имеет вид, схематически показанный на рис. 1.5. Отличительной чертой дельта-слоев является возможность получения в них очень высокой концентрации размерно-квантованных носителей (до значений порядка 10^{14} см$^{-2}$), заметно большей, чем в других описанных структурах. Однако подвижность носителей в них сравнительно невелика за счет рассеяния на большом количестве примесных ионов, лежащих непосредственно в плоскости слоя.

1.4. Структуры с одномерным электронным газом (квантовые нити)

Получение квантовых структур с эффективной размерностью меньше двух (квантовых нитей и точек) является более сложной задачей, чем технология двумерных систем, описанных в предыдущем разделе. Наиболее распространенным способом ее решения является субмикронная литография. При этом исходным объектом является структура с двумерным газом, чаще всего одиночная гетероструктура, показанная на рис. 1.3. Она подвергается литографической процедуре, в ходе которой движение электронов ограничивается еще в одном из направлений. Для этого могут быть использованы два различных подхода [6].

Наиболее очевидный из них — это непосредственное «вырезание» узкой полоски с помощью литографической техники (рис. 1.6, *а*). При этом для получения электронных нитей шириной в десятки

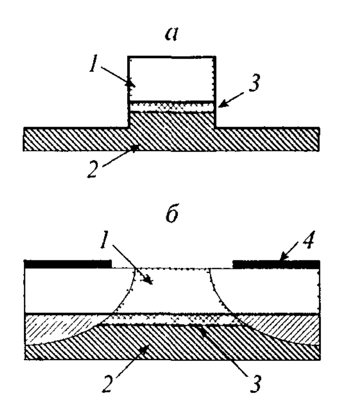


Рис. 1.6. Гетероструктуры с квантовыми нитями, полученные с помощью субмикронной литографии за счет вытравливания узкой полоски из самой структуры (*а*) или щели в затворе Шоттки (*б*).

1 — AlGaAs; *2* — GaAs; *3* — электронный газ; *4* — металлический затвор.

нанометров, где квантование энергии электронов будет заметным, не обязательно делать полоски столь малой ширины, что представляет собой нелегкую технологическую задачу. Дело в том, что на боковых гранях вытравленной полоски, как и на свободной поверхности полупроводника, образуются поверхностные состояния, создающие, как правило, слой обеднения. Этот слой вызывает дополнительное сужение проводящего канала, в результате чего квантовые эффекты можно наблюдать и в полосках большей ширины — порядка десятой доли микрона.

Можно поступить и иначе. Поверхности полупроводниковой структуры с двумерным газом покрывают металлом, создающим с полупроводником контакт Шоттки и имеющим узкую щель (рис. 1.6, *б*). Если гетерограница находится достаточно близко от поверхности, то двумерные электроны будут отсутствовать всюду, кроме узкой области под щелью. Такой тип одномерной структуры обладает дополнительным преимуществом: меняя напряжение на металлическом затворе, мы можем управлять эффективной шириной квантовой нити и концентрацией носителей в ней.

Существуют и другие методы создания квантовых нитей. Способ, изложенный в работе [7], использовался для создания лазеров на структурах с квантовыми нитями. На подложке широкозонного полупроводника (например, AlGaAs) была сделана узкая канавка (или серия параллельных узких канавок) треугольного сечения (рис. 1.7). При эпитаксиальном выращивании на такой подложке квантовой ямы, т. е. тонкого слоя узкозонного полупроводника (GaAs), а затем снова широкозонного материала, толщина слоя GaAs на плоской части подложки (a_0) будет меньше, чем в канавке (a). Поэтому, в согласии с (1.3), размерно-квантованные энергети-

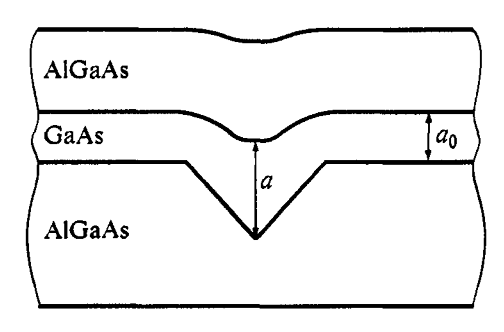


Рис. 1.7. Гетероструктура на подложке с узкой канавкой.

ческие уровни внутри канавки расположены ниже, чем в остальной части слоя GaAs. Это означает, что носители заряда (как электроны, так и дырки) при невысокой их концентрации будут стремиться локализоваться на квантовых уровнях в пределах канавки, образуя вдоль нее квантовую нить.

1.5. Структуры с нуль-мерным электронным газом (квантовые точки)

Принципы создания квантовых нитей, описанные в предыдущем разделе, могут быть также применены и для создания квантовых точек. Для этого фактически требуется лишь изменить картинку, вытравливаемую на двумерной структуре с помощью субмикронной литографии. В случае квантовых нитей она представляла собой систему узких полосок. В случае квантовых точек речь идет об отдельных кружках (или иных изолированных плоских фигурах) субмикронного размера, которые либо оставляются нетронутыми, если используется метод прямого вытравливания, как на рис. 1.6, *а*, либо, наоборот, вытравливаются в металлическом покрытии для структур с контактом Шоттки, показанных на рис. 1.6, *б*.

Следует заметить, что методы субмикронной литографии достаточно дороги и имеют естественные ограничения по размерам, не позволяя получать структуры с предельно малым боковым ограничением, таким же как в направлении роста. Поэтому в технологии квантовых нитей и особенно квантовых точек усиленно ведется поиск альтернативных методов получения таких структур. Особенно привлекательной выглядит возможность использования эффектов самоорганизации, где наноструктуры определенных размеров формируются сами, под влиянием внутренних сил, действующих в процессе роста. Известны два основных способа формирования квантовых точек подобным образом.

Первый способ состоит в выращивании полупроводниковых нанокристаллов с характерными размерами в единицы и десятки нано-

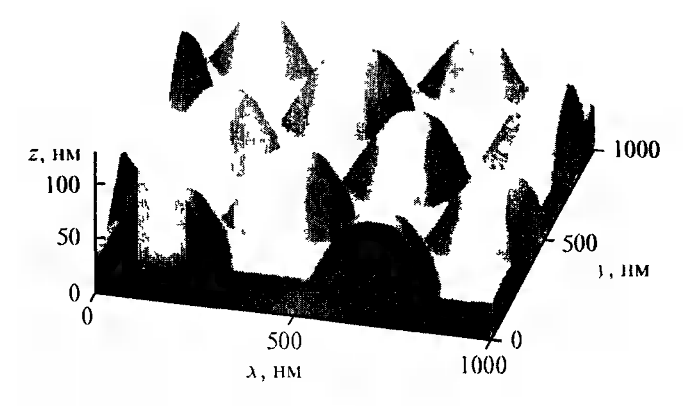


Рис. 1.8. Система самоупорядоченных квантовых точек InAs на поверхности GaAs.

метров. Существует ряд технологических приемов, позволяющих этого добиться [8]. Например, при выращивании полупроводников из пересыщенного раствора в стеклянной матрице процесс формирования зародышей будет определяться диффузией полупроводниковых атомов к центрам роста. Вначале зародыши растут монотонно со временем, но затем, когда пересыщение падает, начинается перераспределение атомов между различными зародышами, приводящее к выравниванию их размеров и формированию достаточно узкой функции их распределения [9]. В результате можно получить большое количество почти одинаковых нанокристаллов с размерами, требуемыми для наблюдения квантовых размерных эффектов, «вмороженных» в стеклянную матрицу. Возможно получать нанокристаллы и в иных прозрачных твердых и жидких средах. Наиболее часто подобным методом получают нанокристаллы полупроводниковых соединений II—VI (CdS, CdSe и др.).

Суть второго способа заключается в формировании слоя квантовых точек при гетероэпитаксии полупроводников с большим рассогласованием параметров решетки. Наиболее распространенной системой здесь является пара GaAs–InAs, для которой параметры решетки отличаются на 7 %. С точки зрения эпитаксии это очень большое рассогласование, и при выращивании достаточно толстого слоя InAs на подложке GaAs мы получили бы сильно дефектную гетерограницу с большой плотностью дислокаций несоответствия. Однако если толщина слоя составляет лишь несколько периодов решетки, энергетически более выгодно деформировать слой, сделав период его решетки таким же, как у подложки, нежели образовывать дислокации несоответствия. Свойства структур с такими напряженными слоями еще будут обсуждаться в этой книге. Сейчас же для нас важно другое обстоятельство. На начальной стадии роста подобной структуры, когда количество атомов InAs

еще недостаточно для полного покрытия подложки, они образуют на подложке отдельные островки. Совместное влияние упругих напряжений на гетерогранице и сил поверхностного натяжения на внешних поверхностях островков приводит к тому, что все островки имеют одинаковые размеры, обычно в несколько нанометров, и образуют периодическую решетку на поверхности подложки. Поскольку ширина запрещенной зоны в InAs значительно меньше, чем в GaAs, то эти островки представляют собой трехмерные потенциальные ямы для носителей, или, иными словами, квантовые точки. На рис. 1.8 показано изображение самоупорядоченной структуры с квантовыми точками InAs на поверхности GaAs, полученное с помощью атомарно-силового микроскопа.

1.6. Структуры с вертикальным переносом

Физика размерного квантования и энергетический спектр соответствующих наноструктур обсуждались в разделе 1.1 на примере одиночных пленок, нитей и точек. Реальные экспериментальные образцы (в особенности квантовые точки) в большинстве случаев содержат большое количество одинаковых или почти одинаковых квантовых объектов. Как правило, это не меняет физической картины, поскольку вклады от всех объектов просто суммируются. Ситуация, однако, резко меняется, если отдельные слои, нити или точки находятся столь близко друг к другу, что носители заряда могут туннелировать между ними.

Рассмотрим систему параллельных квантовых ям, показанную на рис. 1.9, с очень тонкими (порядка единиц нанометров) широкозонными разделяющими слоями. При этом ямы уже не являются независимыми и могут обмениваться электронами за счет туннелирования через широкозонный слой. Подобные структуры принято называть структурами с вертикальным переносом. Рис. 1.9 отвечает системе квантовых ям, но, подвергнув эту систему литографической процедуре (см. раздел 1.4), мы можем иметь систему квантовых нитей или точек, между которыми возможен вертикальный пе-

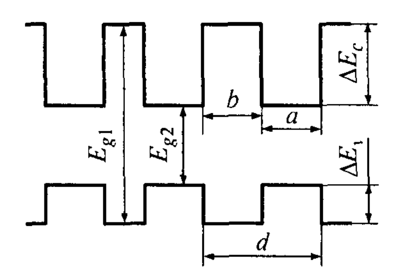


Рис. 1.9. Зонная диаграмма сверхрешетки.

ренос. Подобные структуры служат основой для ряда приборов наноэлектроники, обсуждаемых в главе 8, таких, например, как резонансно-туннельный диод или одноэлектронный транзистор.

Если число параллельных слоев в структурах с вертикальным переносом велико (как минимум, несколько десятков), мы имеем искусственную периодическую структуру, или сверхрешетку [10, 11]. Наиболее важным свойством сверхрешеток, определяющим все их уникальные физические свойства, обсуждаемые в главе 5, является видоизменение их энергетического спектра по сравнению со спектром одиночной квантовой ямы (1.4).

Из рис. 1.9 видно, что на электроны и дырки в сверхрешетке действует дополнительный прямоугольный потенциал $V(z)$, связанный с разрывами зон на гетерограницах. Этот потенциал является периодическим, как и потенциал кристаллической решетки, и к нему применимы все основные выводы о свойствах уравнения Шредингера с периодическим потенциалом, излагаемые в любом курсе физики твердого тела. Напомним важнейшие из них.

Движение носителей вдоль оси z (ось сверхрешетки) может быть описано с помощью квазиимпульса p_z, причем энергия является периодической функцией p_z с периодом $2\pi\hbar/(a + b)$. Энергетический спектр носит зонный характер и представляет собой чередование разрешенных $E_n(p_z)$ и запрещенных зон. Эти зоны есть результат дробления исходной зоны проводимости (для электронов) и валентной зоны (для дырок), поэтому их принято называть минизонами.

Наибольший интерес представляют минизоны, образованные из нижних уровней квантовой ямы, где сконцентрирована основная масса носителей. Их ширина определяется вероятностью туннелирования из ямы в яму и потому невелика. Такие минизоны хорошо описываются в приближении сильной связи, которое приводит к закону дисперсии

$$E_n(p_z) = E_n + \Delta_n \cos[p_z(a+b)/\hbar]. \qquad (1.10)$$

Между энергетическими спектрами сверхрешетки и обычной кристаллической решетки существуют большие различия.

Во-первых, зонным спектром (1.10) характеризуется лишь движение вдоль оси сверхрешетки z. В плоскости слоев носители движутся как свободные и полный спектр сверхрешетки резко анизотропен, так же как в рассмотренных выше двумерных системах (1.4):

$$E = E_n + \Delta_n \cos[p_z(a+b)/\hbar] + \frac{p_x^2 + p_y^2}{2m}. \qquad (1.11)$$

Во-вторых, благодаря тому, что период сверхрешетки намного больше, чем параметр решетки кристалла, характерный масштаб

энергий и импульсов в спектре $E_n(p_z)$ оказывается значительно меньшим. Характерные ширины минизон $2\Delta_n$ измеряются десятыми или сотыми долями электронвольта, что сравнимо с тепловой энергией электрона. Поэтому движение по минизоне нельзя описать постоянной эффективной массой. Динамика в направлении оси сверхрешетки носит сложный характер, что служит причиной ряда необычных эффектов, рассматриваемых в главе 5.

При увеличении толщины барьерных слоев b их туннельная прозрачность уменьшается и минизоны сужаются, превращаясь в дискретные уровни E_n отдельных слоев. В противоположном случае $b \to 0$ сужаются все запрещенные минизоны и мы приходим к случаю однородного полупроводника. Поэтому формально сверхрешетки можно рассматривать как промежуточный случай между дву- и трехмерными электронными системами.

Существуют также непериодические квантовые сверхрешетки, в которых параметры чередующихся слоев (толщина барьера или ширина ямы) меняются по определенному закону либо совершенно случайным образом. Энергетический спектр таких сверхрешеток существенно отличается от спектра периодических сверхрешеток, так, например, в сверхрешетках со случайным чередованием полностью отсутствуют минизоны. Среди непериодических сверхрешеток выделяются квазипериодические сверхрешетки, у которых решетка образуется путем чередования двух структурных элементов А и В, каждый из которых состоит из барьера и ямы. Структурные элементы отличаются либо шириной ямы, либо шириной (или высотой) барьера. В сверхрешетке Фибоначчи [12] последовательное увеличение количества слоев в решетке происходит по закону Фибоначчи и решетки по мере роста количества слоев образуют иерархическую структуру S_n, в которой каждый последующий член ряда образуется как последовательное соединение двух предыдущих: $S_n = S_{n-2} + S_{n-1}$. Так, например, S_1 = BA, S_2 = ABA, S_3 = BAABA,..., S_∞ = ABAABABAABA... Существуют и другие законы образования непериодических сверхрешеток. Их особенностью является фрактальная структура энергетического спектра и самоподобие спектра плотности состояний. Это выражается в том, что в решетках с достаточно большим количеством слоев структура спектра плотности состояний в целом оказывается такой же, как и в небольшом интервале энергий. При этом в спектре плотности состояний проявляется чередование широких и узких ступеней. По своим свойствам квазипериодические сверхрешетки занимают промежуточное положение между периодическими и совершенно неупорядоченными сверхрешетками.

Глава 2

ТЕХНОЛОГИЯ КВАНТОВО-РАЗМЕРНЫХ СТРУКТУР

2.1. Проблемы технологии квантово-размерных структур

Прогресс в экспериментальном изучении физических свойств квантово-размерных структур тесно связан с развитием технологии их получения и в значительной мере им и определяется.

Изготовление квантово-размерных структур в большинстве случаев требует создания полупроводниковых гетеропереходов с необходимыми свойствами. Для этого прежде всего необходимо подобрать подходящую пару полупроводниковых материалов. Выбор этих материалов зависит от типа структуры, которую необходимо получить. Для создания структур с двумерным электронным газом на основе гетеропереходов основным условием является требование равенства постоянных решетки у обоих полупроводников. Нарушение этого условия может привести к образованию высокой плотности дислокаций несоответствия вблизи гетерограницы, что резко ухудшает свойства переходов и делает невозможным наблюдение эффектов размерного квантования.[1]

Гетеропереход $GaAs–Al_xGa_{1-x}As$ наиболее часто используют для изготовления квантово-размерных структур. Эти материалы обладают хорошим согласием решеток при любом составе твердого раствора и потому позволяют создавать гетеропереходы с различными разрывами зон на границе. Среди полупроводников класса $A^{III}B^{V}$ есть еще несколько идеальных гетеропар, например $InP–In_{0.53}Ga_{0.47}As$, но они обладают согласием решеток лишь при одном фиксированном составе и потому не позволяют варьировать ΔE_c и ΔE_v. Другие классы полупроводников реже используются для изготовления квантовых гетероструктур.

Одно из главных требований к технологии изготовления квантовых гетероструктур связано с необходимостью получения очень

---

[1] Сверхрешетки с достаточно тонким слоем можно выращивать и при наличии рассогласования решеток. В тонких слоях дислокации несоответствия не образуются, а различия в постоянных решетки вызывают упругую деформацию слоев, изменяющую энергетический спектр системы.

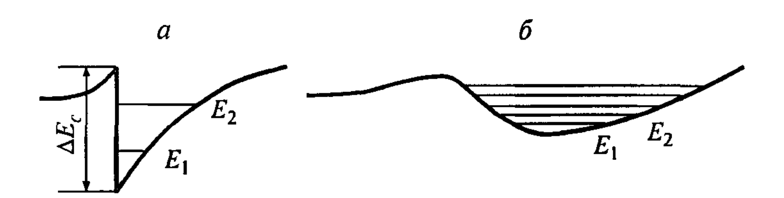


Рис. 2.1. Потенциальный профиль края зоны проводимости для резкого (*а*) и плавного (*б*) гетеропереходов.

резких гетеропереходов с переходным слоем промежуточного состава, имеющим толщину всего в несколько постоянных решетки. Как видно из рис. 2.1, для плавного перехода образуется потенциальная яма большой ширины, что уменьшает расстояние между уровнями E_n и затрудняет наблюдение квантовых размерных эффектов.

Другим важным требованием к технологии является возможность получения сложных профилей состава и легирования. При изготовлении сверхрешеток необходимо строго периодически менять состав растущего слоя, причем период изменения может составлять лишь несколько десятков ангстрем. При выращивании δ-слоев (см. главу 1) и структур с модулированным легированием (см. главу 5) необходимо со столь же высокой точностью менять концентрацию легирующих примесей.

Далеко не всякая эпитаксиальная технология удовлетворяет приведенным требованиям. Плохо пригодными, в частности, оказываются наиболее распространенные методики газотранспортной и жидкостной эпитаксии. Получение сложного профиля состава и легирования наталкивается на ряд технических трудностей, а сравнительно высокая температура роста вызывает диффузионное размывание гетерограницы, не позволяя получать очень резкие границы.

В настоящее время наиболее часто для изготовления качественных гетероструктур применяют метод молекулярно-лучевой эпитаксии [13]. Другим возможным способом изготовления гетороструктур является газовая эпитаксия из металлоорганических соединений [11]. Она не требует столь сложного и дорогого оборудования, обладает меньшими возможностями контроля и управления, но тем не менее позволяет растить гетеропереходы требуемой резкости и структуры достаточно сложного профиля, включая сверхрешетки, квантовые нити и точки.

Получение одно- и нуль-мерных структур возможно путем выделения их с помощью литографии. Однако это требует применения особых методов литографии — нанолитографии, поскольку столь малые структуры трудно приготовить при помощи стандартных методов.

Значительные успехи были достигнуты в изготовлении квантовых точек и квантовых нитей с использованием эффектов самоорганизации упорядоченных наноструктур на полупроводниковых подложках. Этот метод предъявляет специфические требования к материалам гетеропары, которые будут обсуждаться в разделе 2.5.

2.2. Метод молекулярно-лучевой эпитаксии

Молекулярно-лучевая эпитаксия (МЛЭ) [11, 13] представляет собой усовершенствованную разновидность методики термического напыления в условиях сверхвысокого вакуума. Давление остаточных газов в вакуумной камере поддерживается ниже $1 \cdot 10^{-8}$ Па ($\sim 10^{-10}$ мм рт. ст.).

Потоки атомов или молекул образуются за счет испарения жидких или сублимации твердых материалов, которые располагаются в источнике — эффузионной ячейке. Эффузионная ячейка представляет собой цилиндрический либо конический тигель диаметром 1—2 см и длиной 5—10 см. На выходе ячейка имеет круглое отверстие — диафрагму диаметром 5—8 мм. Для изготовления тигля используют пиролитический графит высокой чистоты либо нитрид бора BN.

Потоки атомов (или молекул) необходимых элементов направляются на подложку и осаждаются там с образованием вещества требуемого состава. Схематическое изображение основных узлов установки МЛЭ приведено на рис. 2.2. Количество эффузионных ячеек зависит от состава пленки и наличия легирующих примесей. Для выращивания элементарных полупроводников (Si, Ge) требу-

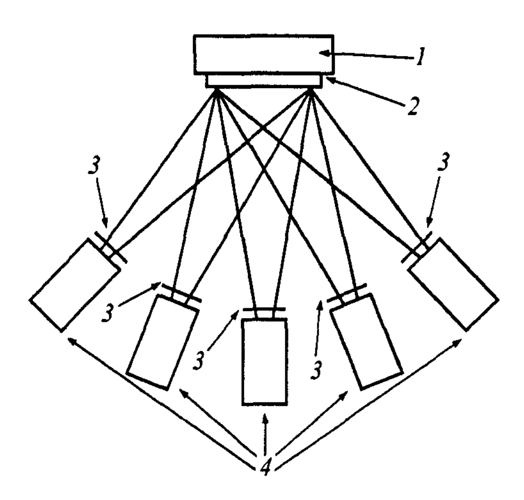


Рис. 2.2. Схематическое изображение основных узлов установки МЛЭ.
1 — подложка; *2* — растущая пленка; *3* — заслонки; *4* — эффузионные ячейки.

ется один источник основного материала и источники легирующей примеси *n*- и *p*-типа. В случае сложных полупроводников (двойных, тройных соединений) требуется отдельный источник для испарения каждого компонента пленки (например, для $Al_xGa_{1-x}As$ требуются отдельные источники Al, Ga, As).[2] Температура эффузионной ячейки определяет величину потока частиц, поступающих на подложку, и тщательно контролируется. Управление составом основного материала и легирующих примесей осуществляется с помощью заслонок, перекрывающих тот или иной поток. Если в ходе выращивания структуры требуется резко менять концентрацию одной и той же примеси, то используют несколько источников этой примеси при разной температуре эффузионной ячейки. Однородность состава пленки по площади и ее кристаллическая структура определяются однородностью молекулярных пучков. В некоторых случаях для повышения однородности подложка с растущей пленкой постоянно вращается.

Наличие высокого вакуума в камере роста полупроводниковой пленки при МЛЭ позволяет использовать различные методы контроля для определения параметров растущей пленки. В различных сочетаниях в установках МЛЭ использовались методы отражательной дифракции медленных электронов, электронная оже-спектроскопия, масс-спектроскопия вторичных ионов, рентгеновская электронная спектроскопия, эллипсометрия и другие оптические методы.

Дифракция медленных электронов является чувствительным методом исследования поверхностной структуры, микроструктуры и степени гладкости. Она используется также для контроля за изменением структуры верхних атомных слоев и для наблюдения топографии поверхности (особенно на ранних стадиях роста). Метод электронной оже-спектроскопии используется для изучения химического состава поверхности подложки, определения химического состава растущего слоя, наблюдения за обогащением поверхности примесями (при условии концентрации примеси более 1% от числа поверхностных атомов), но метод недостаточно чувствителен для изучения типичных уровней легирования в полупроводниковых приборах порядка одной миллионной. Масс-спектроскопия вторичных ионов служит мощным методом определения химического состава внешних атомных слоев твердого тела. Он обладает более высокой чувствительностью, чем оже-спектроскопия. Эллипсометрия позволяет контролировать толщину пленок, отличных по составу от подложки. Другие оптические методы обеспечивают контроль за интенсивностью молекулярных пучков.

---

[2] Материалы $A^{IV}B^{VI}$ можно получить путем испарения полупроводника двойного состава из одной эффузионной ячейки.

В целом установка МЛЭ является сложным устройством со множеством контролируемых параметров. Поэтому современные установки МЛЭ управляются обычно мощными компьютерами.

Выбор подложки для нанесения пленки определяется типом квантово-размерной структуры, которую необходимо получить. Эпитаксиальное наращивание пленки для получения гетероструктуры с двумерным электронным газом или сверхрешетки требует согласования постоянных решетки для предотвращения образования дислокаций несоответствия. Создание же нуль-мерных структур за счет эффектов самоорганизации (см. раздел 2.5), наоборот, предполагает наличие рассогласования решеток для появления механических напряжений. Выращивание высококачественных эпитаксиальных слоев методом МЛЭ требует тщательности в подготовке подложек, поскольку в методе МЛЭ, как правило, не используется очистка поверхности в самой камере роста, за исключением удаления окисных слоев.

Эпитаксиальный рост слоев полупроводниковых соединений включает ряд последовательных событий, важнейшими из которых являются:

1) адсорбция составляющих вещество атомов и молекул;

2) миграция и диссоциация адсорбированных частиц;

3) пристраивание составляющих атомов к подложке, приводящее к зародышеобразованию и росту слоя.

Растущая тонкая пленка имеет кристаллографическую структуру, определяемую подложкой.

Образование пленки связано с фазовым переходом из сильно неравновесной паровой фазы в твердую фазу тонкой пленки. Парциальное давление в пучках компонент в несколько раз превосходит равновесное с подложкой давление пара. Это означает, что процесс осаждения в целом является неравновесным. Поэтому наряду с термодинамическими параметрами в процессе роста важную роль играют такие кинетические параметры, как способность атомов двигаться по поверхности (поверхностная диффузия), среднее время пребывания их на поверхности и поток атомов и молекул на растущую поверхность твердого тела.

Скорость поступления вещества на подложку определяется потоком (интенсивностью пучка) атомов или молекул. При МЛЭ полупроводниковых соединений типичные значения потоков составляют 10^{14}—10^{16} атом/(см$^2 \cdot$ с). Атомы, попадающие на подложку, адсорбируются на поверхности. На первом этапе это представляет собой адсорбцию за счет слабых ван-дер-ваальсовых и (или) электростатических сил — этап физсорбции. На втором этапе молекулы вещества переходят в хемосорбированное состояние, при котором происходит электронный перенос, т. е. химическая реакция между атомами поверхности и атомами вновь поступившего вещества.

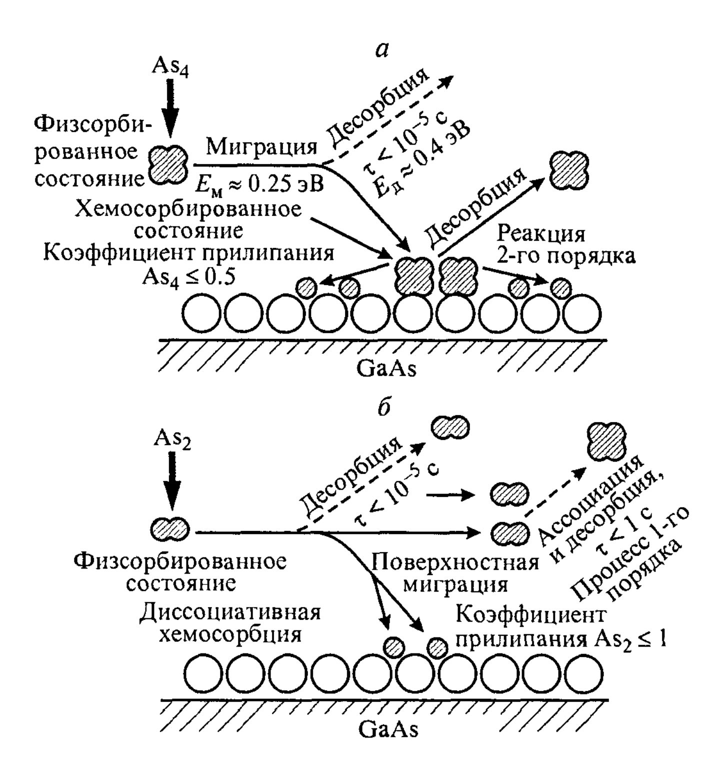


Рис. 2.3. Модель механизма роста GaAs из молекулярных пучков Ga и As_4 (*а*) и пучков Ga и As_2 (*б*).

Энергия связи при химической адсорбции больше, чем при физической.

Прежде чем встроиться в решетку, атом совершает в среднем 10^6 прыжков по поверхности. Такое число прыжков дает массу возможностей для выбора положения с наименьшей свободной энергией. Для полупроводниковых соединений процессы роста в значительной мере определяются химией поверхностных реакций. Рассмотрим их на примере наиболее исследованного материала — арсенида галлия GaAs.

Для выращивания пленок GaAs с помощью МЛЭ используются две методики. Они различаются видом молекул мышьяка, которые поступают в потоке на поверхность роста. В одной из них поток состоит из тетрамерных молекул мышьяка (As_4), а в другой — из димерных молекул (As_2). Поток Ga всегда состоит из отдельных атомов. Молекулярный пучок As_4 создается при загрузке эффузионной ячейки чистым мышьяком, а для создания пучка As_2 в ячейку помещают GaAs либо предварительно пропускают поток As_4 через высокотемпературную зону, где происходит разложение тетрамерных молекул As_4 и превращение его в димерные молекулы As_2.

Участие молекул As_4 в процессе роста выглядит следующим образом (рис. 2.3, *а*): в хемосорбированном состоянии происходит парное взаимодействие молекул As_4, адсорбированных на соседних узлах Ga. Из любых двух молекул As_4 четыре атома As встраиваются в решетку GaAs, а другие четыре десорбируются в виде молекулы As_4. Скорость десорбции в этом случае является показателем скорости роста пленки GaAs. Если поверхностная концентрация As_4 мала по сравнению с числом свободных узлов Ga, то скорость процесса определяется вероятностью столкновения между молекулами As_4 и приводит к квадратичной зависимости скорости десорбции от полного потока As_4, падающего на поверхность. С ростом поверхностной концентрации As_4 растет вероятность того, что пришедшая из падающего потока молекула найдет ближайшие узлы занятыми. При этом скорость десорбции становится пропорциональной числу подводимых молекул. В этом случае часть одиночных мест остается не занятой мышьяком и в растущей структуре появляются вакансии As, которые вызывают появление глубоких уровней в запрещенной зоне.

Взаимодействие атомов Ga и молекул As_2, поступающих в потоке на поверхность, представляет собой простую реакцию диссоциативной хемосорбции на поверхностных атомах Ga (рис. 2.3, *б*). Диссоциация адсорбированных молекул As_2 возможна лишь тогда, когда в ходе миграции по поверхности они наталкиваются на пару свободных узлов Ga. В отсутствие на поверхности свободных адсорбированных атомов Ga молекулы As_2 проводят на поверхности вполне определенное время, а затем испаряются (десорбируются) с поверхности. В условиях такой кинетики стехиометрический GaAs может расти лишь в случае, когда поток молекул As_2 поставляет на поверхность больше атомов As, чем поток атомов Ga, при этом любой избыток As_2 будет десорбироваться.

Описанные модели роста справедливы при температуре роста ниже 600 К. При более высоких температурах роста происходит некоторая диссоциация GaAs, молекулы As_2 уходят с поверхности и происходит обогащение поверхности адсорбированными атомами Ga. Образуется так называемая Ga-стабилизированная структура. В предельном случае на поверхности подложки может образоваться слой чистого Ga. Уходящий с поверхности мышьяк может восполняться при диссоциации падающих пучков As_2 или As_4. Если скорость подвода мышьяка превосходит скорость его ухода, то поверхность обогащается атомами мышьяка и возникает As-стабилизированная структура. Таким образом, отношение потоков мышьяка и галлия, которое требуется для поддержания определенной поверхностной структуры, зависит от температуры подложки и абсолютной величины потоков. Скорость роста кристаллических пленок GaAs при МЛЭ полностью определяется интенсивностью

пучка галлия, падающего на поверхность, в то время как процесс формирования из пучков элементов пленки соединения определяется адсорбцией мышьяка.

Данная модель роста справедлива также для AlAs, InP, большого числа других полупроводниковых соединений $A^{III}B^{V}$ и с незначительными изменениями для тройных соединений $A_x^{III}B_{1-x}^{III}C^{V}$. Пленки $Al_xGa_{1-x}As$ с высокой степенью кристаллического совершенства могут быть выращены просто путем одновременного направления на подложку пучков Ga, Al и избыточного количества As_4. Поверхностный состав твердого раствора $A_x^{III}B_{1-x}^{III}C^{V}$ определяется отношением потоков лишь элементов III группы, если рост идет при температуре ниже границы термической устойчивости менее стойкого соединения $A^{III}B^{V}$ (в случае $Al_xGa_{1-x}As$ термически менее стабильным является GaAs по сравнению с AlAs). При более высоких температурах возможны также сегрегация атомов III группы и обогащение поверхности одним из них.

Введение примесей при МЛЭ обычно осуществляется с помощью дополнительных эффузионных ячеек, содержащих необходимый легирующий элемент. Многие примеси, используемые при других формах эпитаксии, оказываются неподходящими для МЛЭ. В процессе МЛЭ успешно используются лишь те легирующие материалы, атомы которых быстрее связываются с поверхностью, чем испаряются с нее (т. е. материалов, у которых отношение $N_{ад}/N_{полн}$[3] близко либо равно единице, где $N_{ад}$ — число атомов, адсорбированных поверхностью, а $N_{полн}$ — полное число подлетающих частиц). Результирующие электрические и оптические свойства пленок зависят от соотношения потоков основных элементов и потока примеси, температуры подложки, а также характера реального встраивания примеси в решетку и степени их электрической активности. Наиболее важными проблемами для большинства примесей являются накопление их на поверхности, десорбция, образование на поверхности соединений, образование комплексов и автокомпенсация.

Наиболее изучены к настоящему времени примеси в GaAs. Акцепторными свойствами, наиболее приемлемыми в GaAs, обладают бериллий и магний. Германий обладает амфотерными свойствами, и его практическое применение ограничивается необходимостью тщательного выбора и поддержания температуры подложки и соотношения потоков мышьяка и галлия. Наиболее удачной легирующей примесью является бериллий. Он позволяет получать *p*-GaAs с концентрацией до $2 \cdot 10^{20}$ $см^{-3}$. При введении Be не возникает сегрегации или аномальной объемной диффузии примеси. Концентрация активных акцепторов оказывается прямо пропор-

---

[3] Это отношение называется коэффициентом прилипания.

циональной температуре эффузионной ячейки, т. е. давлению пара Be. Каждый атом Be создает активный акцептор с уровнем, лежащим на 30 мэВ выше края валентной зоны.

Донорными свойствами в GaAs обладают кремний, олово, сера и селен. Германий, как амфотерная примесь, при определенных условиях тоже позволяет получить *n*-тип проводимости. Удобной примесью является олово. Концентрация активных доноров прямо пропорциональна потоку Sn в молекулярном пучке и может быть доведена до $5 \cdot 10^{19}$ см$^{-3}$. Однако его существенным недостатком является скапливание при МЛЭ у растущей поверхности GaAs. По этой причине наиболее предпочтительной примесью донорного типа является кремний. Он обеспечивает концентрацию дырок в GaAs до $1.3 \cdot 10^{19}$ см$^{-3}$.

В заключение отметим основные характеристики метода МЛЭ:

1) малая скорость роста, порядка 1 мкм/ч;

2) относительно низкая температура роста (температура подложки);

3) возможность резкого прерывания и возобновления роста за счет использования механических заслонок вблизи эффузионных ячеек для всех компонентов;

4) возможность введения различных парообразных компонентов для изменения состава слоя и управления концентрацией примесей путем введения дополнительных источников, создающих требуемые пучки молекул или атомов;

5) наличие атомно-гладкой поверхности растущего кристалла при эпитаксиальном росте;

6) возможность анализа и контроля в ходе роста.

2.3. Газофазная эпитаксия из металлоорганических соединений

«Металлоорганика» — широкий класс веществ, содержащих металл-углеродные (органометаллические соединения) или металл-кислород-углеродные связи (алкоксиды) и координационные соединения металлов с органическими молекулами. При комнатной температуре они в большинстве являются жидкостями, хотя некоторые из них даже при высокой температуре остаются твердыми. Эти вещества обычно имеют высокое давление паров и легко могут быть доставлены в зону реакции путем пропускания газа-носителя, например водорода или гелия, через жидкость или над твердым телом, которые играют роль источника [11, 13].

Кристаллизация в процессе эпитаксии металлоорганических соединений (МОС) осуществляется при пропускании однородной газовой смеси реагентов с газами-носителями над нагретой подлож-

кой в реакторе с холодными стенками. Используемые для роста полупроводниковых пленок гидриды при комнатной температуре являются газами и обычно применяются в качестве небольших добавок к H_2. Эти металлоорганические и гидридные компоненты смешиваются в газовой фазе и пиролизуются в потоке H_2. В результате реакции пиролиза, когда газообразные соединения разлагаются на компоненты на горячей поверхности с образованием твердого осадка, образуется стабильное твердое полупроводниковое соединение. Температура пиролиза составляет 600—800 °С. Нагрев подложки и растущей пленки обычно создается мощным радиочастотным генератором с частотой порядка 450 кГц. Пиролиз происходит в открытом реакторе при атмосферном или пониженном (~70 мм рт. ст.) давлении. Снижение давления газовой смеси при выращивании соединений позволяет эффективно управлять как градиентом изменения состава примесей, так и градиентом изменения состава основных компонент. Пониженное давление газовой смеси также ограничивает протекание паразитных реакций в газовой смеси. При пониженном давлении процесс выращивания ведется при значительно больших скоростях газового потока. Увеличение скорости газового потока в принципе позволяет получить более однородные слои. При пониженном давлении имеется возможность увеличивать скорость движения газов без увеличения потока массы холодного газа. Увеличение скорости потока в условиях атмосферного давления привело бы к охлаждению подложки. Существуют реакторы, в которых поток газовой смеси направлен перпендикулярно поверхности подложки (чаще это реакторы вертикального типа), и реакторы, в которых поток газовой смеси движется вдоль поверхности либо под небольшим углом к ней (обычно это реакторы горизонтального типа).

Вблизи от поверхности роста располагается переходная область, в которой параметры газовой смеси плавно изменяются от значений, характерных для области конвекции, до некоторых значений, характерных для приповерхностного слоя. Эта переходная область называется пограничным слоем. При направлении потока вдоль поверхности толщина его равна примерно 4 мм, а при направлении перпендикулярно подложке 20 мм. В пределах пограничного слоя температура газовой смеси и ее состав зависят от расстояния до поверхности роста. В области конвекции температура газа меньше температуры роста и состав газовой фазы не меняется. Во многих случаях электрические и кристаллографические свойства получающихся слоев зависят от характеристик пограничного слоя. Пример реактора вертикального типа для получения материалов $A^{III}B^{V}$ приведен на рис. 2.4.

Многослойные, многокомпонентные эпитаксиальные структуры могут быть последовательно выращены в едином ростовом цикле.

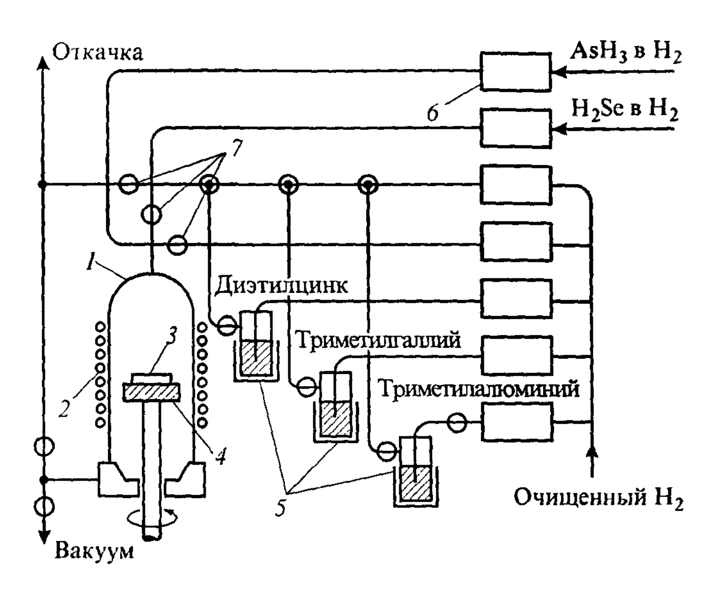


Рис. 2.4. Схематическое изображение вертикального реактора газофазной эпитаксии МОС, используемого для роста легированных эпитаксиальных слоев $Al_xGa_{1-x}As$.

1 — кварцевый реактор; *2* — высокочастотный нагрев; *3* — подложка; *4* — держатель; *5* — термостабилизированные ячейки; *6* — датчики потока; *7* — пневматические клапаны.

К реактору можно подключить несколько металлоорганических и гидридных источников.

Химия технологического процесса определяет скорость роста, качество и чистоту кристалла, его электрические и оптические свойства, возможность управления составом. Для многих полупроводников типа $A^{III}B^{V}$ результирующая реакция записывается в виде

$$A^{III}(CH_3)_3 + B^{V}H_3 \xrightarrow[700\,^{\circ}C]{H_2} A^{III}B^{V} + 3CH_4. \qquad (2.1)$$

В частности, такого типа реакция используется для получения GaAs и AlAs. С помощью аналогичных реакций можно вырастить тонкие пленки других бинарных, тройных и четверных соединений. Например, для роста $Al_xGa_{1-x}As$ обычно используется следующий процесс:

$$(1-x)\,[Ga(CH_3)_3] + x\,[Al(CH_3)_3] + AsH_3 \xrightarrow[700\,^{\circ}C]{H_2} Al_xGa_{1-x}As + 3CH_4. \qquad (2.2)$$

В этом случае состав эпитаксиальной пленки непосредственно определяется отношением парциальных давлений триметилгаллия и триметилалюминия в газовой фазе.

Легирование растущего эпитаксиального слоя осуществляется путем введения в газовый поток соответствующего реагента. Для легирования соединений $A^{III}B^{V}$ примесями p-типа обычно используют органометаллические реагенты: диэтилцинк и бисциклопентадиенилмагний, а для легирования примесями n-типа используют гидриды H_2Se и SiH_4. Органометаллические реагенты легко разлагаются при температуре роста, поэтому поступление примесей в растущий слой ограничивается диффузией через пограничный слой и поверхностной кинетикой. Небольшая доля примеси поступает в растущий слой, а бóльшая часть диффундирует в газовую фазу. При этом уровень легирования не зависит от скорости роста слоя, но экспоненциально уменьшается при увеличении температуры и растет при увеличении парциального давления арсина AsH_3. Эффективность легирования кремнием ограничивается термохимией используемых гидритов. Скорость разложения их на поверхности мала, и большая доля примеси, образуемой на поверхности, входит в растущий слой. В этом случае уровень легирования обратно пропорционален скорости роста и экспоненциально возрастает при увеличении температуры. Уровень легирования кремнием не зависит от парциального давления AsH_3.

Скорость роста определяется скоростью диффузии через пограничный слой компонент, содержащих элементы третьей группы. Для элементов пятой группы имеет место баланс между газовой и твердой фазами. При температурах выше 800 K скорость роста от температуры не зависит, как не зависит скорость роста и от давления арсина, которое, для того чтобы получить гладкие слои высокого качества, должно по меньшей мере на порядок превосходить давление компоненты, содержащей элемент третьей группы. Увеличение скорости газового потока ускоряет рост эпитаксиального слоя. Увеличение парциального давления реагентов в газовом потоке также во всех случаях увеличивает скорость роста.

Рост многослойных структур осуществляется путем изменения газовой атмосферы в реакторе. Скорость, с которой осуществляется такое изменение, зависит от величины потока и геометрии реактора. При больших скоростях в потоке можно производить замену газа достаточно быстро и получать при этом резкие гетеропереходы. Время, необходимое для замены, определяется величиной потока. Поэтому резкость границ оказывается связанной со скоростью роста и, например при росте соединений $A^{III}B^{V}$, зависит от концентрации в газовом потоке через реактор реагента, содержащего элемент третьей группы. Экспериментальные данные по сверхрешетке $GaAs{-}Al_xGa_{1-x}As$ позволяют сделать вывод, что при использовании автоматизированных систем роста удается создавать сверхрешетки с толщиной слоев до 1.5 нм (в отдельных случаях до 0.7 нм), с высоким структурным совершенством. Изменение

состава происходит практически на толщине одного монослоя. Для быстрого изменения состава газовой фазы у поверхности роста требуется оптимизация конструкции реактора для устранения размытия концентрационного фронта в газовом потоке из-за различных диффузионных процессов.

МЛЭ и газофазная эпитаксия из МОС являются хорошими методами выращивания сверхрешеток в системе GaAs–$Al_xGa_{1-x}As$. Технологический процесс газофазной эпитаксии из МОС может быть воспроизведен в большом объеме, и в результате можно получить эпитаксиальные структуры большой площади. При этом могут быть достигнуты высокие скорости роста, удовлетворяющие требованиям промышленности при массовом производстве приборов. МЛЭ, скорее всего, будет доминировать в фундаментальных исследованиях квантово-размерных структур, а газофазная эпитаксия из МОС — в массовом производстве и исследовании приборов.

2.4. Методы нанолитографии

Наиболее очевидным, хотя и не самым простым по исполнению, методом получения одно- и нуль-мерных структур является субмикронная литография [14]. Она позволяет вырезать области, ограниченные по одному или двум направлениям, используя в качестве исходного объекта структуры с двумерным электронным газом. Далеко не все методы литографии, используемые в производстве микроэлектронной аппаратуры, годятся для получения малых по своим размерам объектов. Хотя, как указано в разделе 1.4, область движения можно ограничить и за счет поверхностных областей обеднения, но тем не менее стараются разработать такие приемы изготовления, которые обеспечивали бы получение непосредственно малых областей с ограничением по одному или двум направлениям.

Современные интегральные схемы производят при помощи фотолитографии. Эти же приемы используются для изготовления квантовых нитей и точек. Фотолитография — это процесс, в ходе которого пучок видимого или ультрафиолетового света падает через маску нужной формы на тонкое покрытие из светочувствительного материала — фоторезиста, покрывающее полупроводниковую структуру. Фоторезист — сложная полимерная светочувствительная композиция. Фоторезист, у которого растворимость освещенного участка уменьшается, называется негативным, а фоторезист, растворимость которого после облучения возрастает, — позитивным. После обработки освещенного фоторезиста в составе, удаляющем растворимые участки, образуется рельефное изображение, которое должно быть устойчивым к воздействию техно-

логических факторов, в частности кислот, с помощью которых стравливается полупроводниковая структура.

Изображение на поверхность фоторезиста переносят с помощью оптической проекции маски. Можно использовать также тонкий луч света, положением которого управляет компьютер, и вычерчивать рисунок на поверхности фоторезиста или использовать интерференцию двух лазерных лучей, в результате которой появляется периодическая модулированная картина. Наименьшие размеры отдельных деталей, которые требуются для изготовления квантовых структур, ограничены предельной разрешающей способностью оптических устройств, которая в свою очередь определяется дифракционным критерием Рэлея. В соответствии с ним минимальная ширина линии, получаемой в изображении, пропорциональна длине волны экспонирующего света. Для видимого света (с длиной волны ~0.4 мкм) минимальная ширина линии составляет ~0.25 мкм, что явно недостаточно для изготовления квантовых структур.

Чтобы достигнуть меньших размеров отдельных деталей, литография с видимым светом заменяется на литографию с использованием ультрафиолетового освещения (длина волны 250—350 нм). Достоинством ультрафиолетовой литографии является большая скорость, так как время освещения меньше, чем для видимого света.

Дальнейшее уменьшение размеров отдельных деталей возможно при использовании рентгеновской литографии. Для экспонирования используют низкоэнергетическое излучение 1—10 кэВ. Это соответствует длинам волн 0.4—5 нм. Проблемой использования рентгеновского излучения является то, что для него не существует линз и хороших зеркал, поэтому источник рентгеновского излучения должен быть удален достаточно далеко для уменьшения расходимости лучей. Из-за большой энергии квантов рентгеновского излучения необходимо использовать более толстые маски толщиной 1—10 мкм. Существуют также проблемы с переизлучением материалов самой маски, эмиссией оже-электронов из резиста, что приводит к снижению контраста изображения. Высокая энергия квантов рентгеновского излучения увеличивает возможность повреждения подложки. Достоинствами метода являются малое время экспонирования (около 1 мин) и возможность получения изображения на большой площади. Снизить время экспонирования за счет большей мощности излучения до нескольких секунд можно путем применения синхротронного излучения. С помощью рентгеновской литографии удается получить минимальный размер деталей до 30 нм.

Повысить разрешение можно с использованием электронно-лучевой литографии. Электроны имеют малую длину волны и могут

быть сфокусированы очень точно при помощи электрического поля, управляемого компьютером. Можно использовать также маски. Проблемой в использовании электронного луча является возможность повреждения поверхности из-за столкновений высокоэнергичных электронов с подложкой. Такие повреждения вносят дефекты в полупроводниковую структуру. Для уменьшения повреждения можно снизить ток в луче, но это затрудняет фокусировку луча. Электронная литография позволяет получить размер отдельных деталей 3—25 нм.

Еще большей разрешающей способностью (из-за большей массы частицы длина волны де Бройля меньше) обладает ионно-лучевая литография. Поскольку удельные потери ионов в десятки раз больше, чем у электронов, то резисты в этом случае обладают более высокой чувствительностью. С помощью ионно-лучевой литографии можно получить изображение и без использования резиста, за счет создания участков с радиационными дефектами. В таких местах возрастает скорость травления и окисления. Ионно-лучевая литография позволяет получить профиль рельефа, близкий к прямоугольному.

Рисунок из фоторезиста передается на материал полупроводниковой структуры при вытравливании в ней мест, не защищенных фоторезистом. При влажном химическом травлении возникают проблемы, которые приводят к нарушению рисунка или профиля рельефа структуры. Так, из-за сил поверхностного натяжения и малых размеров элементов рисунка возникают «мостики» из жидкости и травитель не попадает на поверхность полупроводниковой структуры, в результате чего травление не происходит. В анизотропном травителе скорость травления зависит от кристаллографического направления. В результате может происходить подтравливание под фоторезистом или профиль травления становится отличным от прямоугольного.

Эти недостатки отсутствуют при «сухом» травлении. Так, ионное травление разрушает поверхность в результате бомбардировки высокоэнергичными ионами с энергией в несколько килоэлектронвольт. Мощные пучки ионов с малым диаметром пятна можно перемещать по поверхности с помощью отклоняющих электродов. В этом случае получают наноструктуры и без использования резиста. К достоинствам этого метода можно отнести высокую разрешающую способность и отсутствие подтравливания. Однако при ионном травлении происходит нарушение поверхностного слоя и возможно появление поверхностных энергетических состояний. Другой разновидностью «сухого» травления является ионно-плазменное травление. Оно происходит в плазме реактивного газа. В этом случае ионы имеют невысокую энергию и распыление не происходит. Травление представляет собой химическую реакцию, в

результате которой образуются летучие соединения и откачиваются из камеры. В ионную плазму можно добавить химически активный газ и тем самым менять скорость травления за счет химического взаимодействия газа с поверхностью структуры. Ионно-плазменное травление позволяет организовать анизотропное травление кристаллической структуры.

Наряду с непосредственным выделением нитей и точек из области двумерного электронного газа использовался также рост на профилированных подложках. В этом случае методами литографии на полупроводниковой подложке создавали серии параллельных или пересекающихся канавок с минимальными размерами, которые позволяла получить литография, а затем на эту подложку наносили полупроводниковую пленку. В результате отмечалось влияние профилированных подложек на структуру полупроводниковой пленки. Рост пленки первоначально происходит в углах канавок, вытравленных в подложке. Таким путем можно в принципе получить квантовые нити и точки.

Однако методы литографии и селективного травления обладают рядом недостатков. Так, литографическое выделение рисунка с последующим травлением всегда сопровождается неоднородностью формы наноструктур и механическими повреждениями поверхности. Неоднородность формы приводит к образованию ансамблей неоднородных нитей и точек. Кроме того, граница раздела после травления проявляет заметную дефектность.

2.5. Самоорганизация квантовых точек и нитей

Более перспективными по сравнению с традиционными методами с применением литографии оказались методы прямого получения с использованием эффектов спонтанного образования наноструктур. Эти эффекты относятся к широкому классу фундаментальных явлений самоорганизации в конденсированных средах [15—17].

2.5.1. Режимы роста гетероэпитаксиальных структур

Метод получения квантовых одно- и нуль-мерных структур опирается на рассмотренные ранее методы МЛЭ и газофазной эпитаксии из МОС. Различают три режима роста гетероэпитаксиальных структур, схематически изображенных на рис. 2.5:

1) Франка—ван дер Мерве (Frank—van der Merve, FM) — реализуется послойный (двумерный) рост материала В на подложке из материала А, материалы А и В согласованы по постоянной решетки;

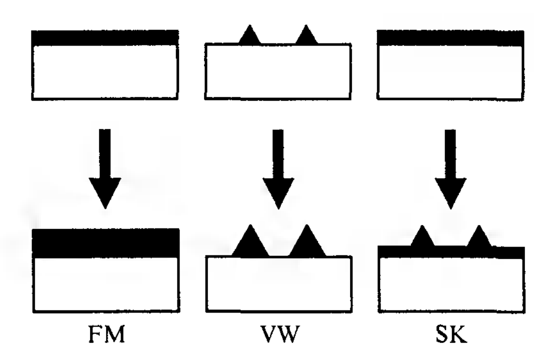


Рис. 2.5. Схематическое изображение трех режимов роста для гетероэпитаксиальных систем: Франка—ван дер Мерве (FM), Фолмера—Вебера (VW) и Странского—Крастанова (SK).

2) Фолмера—Вебера (Volmer—Weber, VW) — имеет место островковый (трехмерный) рост материала В на подложке А;

3) Странского—Крастанова (Stranski—Krastanow, SK) — на первых этапах роста реализуется послойный рост материала В на подложке А с образованием смачивающего слоя, а затем происходит переход к росту трехмерных островков из материала В на покрытой подложке; материалы А и В рассогласованы по постоянной решетки.

В гетероэпитаксиальных системах, согласованных по постоянной решетки, режим роста определяется только соотношением энергий свободных поверхностей материалов А — γ_1, В — γ_2 и энергии границы раздела γ_{12}. Когда поверхностная свободная энергия подложки γ_1 превосходит сумму поверхностной свободной энергии эпитаксиального слоя γ_2 и энергии границы раздела γ_{12}, т. е. $\gamma_2 + \gamma_{12} < \gamma_1$, тогда осаждаемый материал В смачивает поверхности подложки А и возникает режим Франка—ван дер Мерве. Изменение величины $\gamma_2 + \gamma_{12}$ относительно γ_1 приводит к переходу от режима Франка—ван дер Мерве к режиму Фолмера—Вебера. В этом случае материал В не смачивает материал подложки А.

В гетероэпитаксиальной системе при наличии рассогласования по постоянной решетки между осаждаемым материалом В и подложкой А на первых этапах рост может происходить послойно. Однако более толстый слой осажденного материала имеет и бóльшую упругую энергию. Тогда возникает тенденция уменьшить упругую энергию путем образования изолированных островков. В этих островках происходит релаксация упругих напряжений и соответствующее уменьшение упругой энергии. Так возникает режим роста Странского—Крастанова. В конечном итоге на подложке образуются островки, промежутки между которыми заполнены тонким смачивающим слоем.

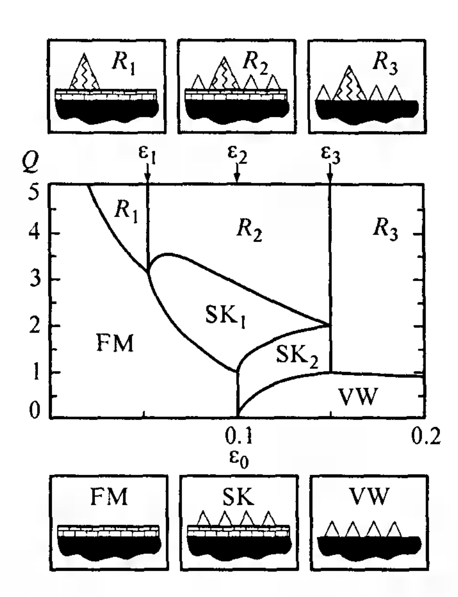


Рис. 2.6. Равновесная фазовая диаграмма гетероэпитаксиальной системы с рассогласованием решеток в осях.

Q — количество осажденного материала; ε_0 — рассогласование решеток.

Термодинамический анализ [18] показал, что на фазовой диаграмме (рис. 2.6), где по одной из осей отложено рассогласование решеток осаждаемого материала и подложки $\varepsilon_0 = \Delta a/a$, а по другой — количество осажденного материала Q (количество монослоев), можно выделить области этих параметров, которые напоминают, а в некоторых областях параметров полностью аналогичны с приведенными выше режимами роста.

В настоящее время известны следующие типы наноструктур, выращиваемых с использованием эффектов самоорганизации:

1) структуры на микроскопически упорядоченных фасетированных поверхностях;

2) упорядоченные массивы трехмерных когерентно напряженных островков в гетероэпитаксиальных рассогласованных системах;

3) периодические структуры плоских доменов (например, островков монослойной высоты);

4) структуры с периодической модуляцией состава в эпитаксиальных пленках твердых растворов полупроводников.

Достоинством способов получения этих наноструктур является то, что структуры появляются как бы сами собой, без применения специальных приемов технологии микроэлектроники и, в частности, без применения литографии.

2.5.2. Рост наноструктур на фасетированных поверхностях

Фасетирование поверхности кристалла представляет собой образование микроскопически упорядоченных «холмов и канавок» на поверхности, которые имеют определенный период повторения. На поверхности образуется система наклонных к исходной поверхности кристаллических граней. Пересечение соседних граней дает в результате или острое кристаллическое ребро, или узкую закругленную область поверхности, профиль такой структуры напоминает зубья пилы (рис. 2.7). Причиной спонтанного фасетирования является зависимость свободной энергии поверхности от ориентации этой поверхности относительно кристаллографических осей. Если плоская поверхность имеет большую удельную поверхностную энергию, то она спонтанно трансформируется в структуру «холмов и канавок». Это уменьшает полную свободную энергию поверхности, несмотря на увеличение ее полной площади. Результирующая холмистая структура определяется минимумом поверхностной свободной энергии.

Атомы на поверхности находятся в ином окружении, чем атомы в объеме, а равновесное расстояние между поверхностными атомами отличается от равновесного расстояния между атомами в объеме. Поэтому поверхность кристалла можно рассматривать как упругонапряженный слой, в котором присутствуют поверхностные напряжения.[4] В результате упругая энергия кристалла имеет как объемную [19], так и поверхностную [20] составляющие и в условиях механического равновесия между объемом и поверхностью возникает отличная от нуля деформация поверхности и приповерхностной области кристалла. Образование холмистой структуры поверхности приводит к капиллярным явлениям, аналогичным лапласову давлению под искривленной поверхностью жидкости. Полную свободную энергию периодически фасетированной поверхности в расчете на единицу площади первоначально плоской поверхности можно записать в виде

$$E = E_{\text{пов}} + E_{\text{реб}} + \Delta E_{\text{упр}}, \tag{2.3}$$

где $E_{\text{пов}}$ — поверхностная свободная энергия наклонных граней (от периода повторения «холмов и канавок» D не зависит); $E_{\text{реб}} = C_1 D^{-1}$ — короткодействующая составляющая энергии ребер; $\Delta E_{\text{упр}} = -C_2 D^{-1}\ln(D/a)$ — упругая энергия релаксации, которая появляется благодаря образованию ребер и скачку тензора собственных поверхностных напряжений на кристаллическом ребре

---

[4] По аналогии с жидкостями собственное поверхностное напряжение кристалла называют также поверхностным натяжением, хотя по физическому смыслу их следует различать.

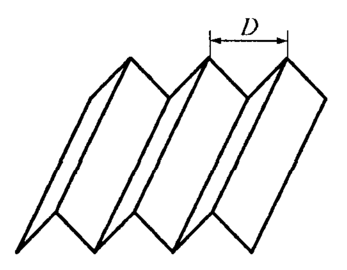


Рис. 2.7. Периодически фасетированная поверхность.

(a — постоянная решетки). Упругая энергия релаксации имеет отрицательный знак и, уменьшая полную свободную энергию поверхности, способствует образованию периодической фасетированной структуры на поверхности с некоторым оптимальным периодом $D_{опт}$.

Фасетированные структуры наблюдались на поверхностях с большими индексами Миллера в GaAs{(311) и (775)}, Si(211), на низкоиндексных поверхностях TaC(110), Ir(110). Важным частным случаем фасетирования поверхности является фасетирование вицинальных поверхностей. Вицинальная поверхность — это поверхность, расположенная под небольшим углом к кристаллической поверхности с малыми индексами Миллера. Обычно такая поверхность является атомно-гладкой. В равновесных условиях (например, после длительного отжига) вицинальная поверхность кристалла состоит из плоских террас, образованных поверхностями с малыми индексами Миллера и разделенных эквидистантными моноатомными или мономолекулярными ступенями. Из-за эффекта «складывания» ступеней (step-bunching), который усиливается по мере увеличения угла разориентации, соседние террасы разделяются равными ступенями высотой в несколько моноатомных слоев (рис. 2.8). Такие ступени и террасы наблюдались на вицинальных поверхностях, рассогласованных на небольшой угол в Si(111), GaAs{(001) и (100)}, Pt(100). Высота ступеней на вицинальной поверхности GaAs(100) составляет от 7 до 15 монослоев. Обычно ступени имеют оптимальный период распространения по одному из направлений вдоль поверхности, а по другому направлению сту-

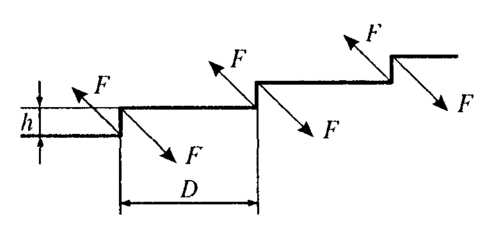


Рис. 2.8. Фасетирование на вицинальной поверхности.

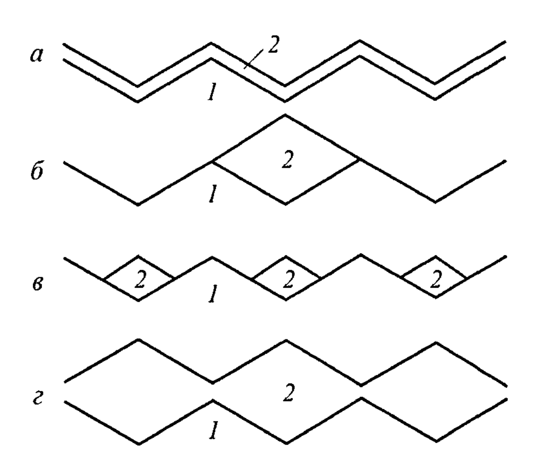


Рис. 2.9. Возможные гетероэпитаксиальные структуры на фасетированной поверхности.

1 — материал подложки; *2* — осажденный материал.

пени отсутствуют. Это создает линейчатую структуру поверхности. Однако при некоторых направлениях разориентации, в частности при разориентации поверхности GaAs(001) в направлении [010], ступени распространяются по двум направлениям: [110] и [$\bar{1}$10]. Пересечение ступеней формирует на поверхности сеть террас, ограниченных со всех четырех сторон.

Периодически фасетированные поверхности дают возможность для прямого получения упорядоченных массивов квантовых проволок, когда рост осаждаемого материала происходит в «канавках», и квантовых точек, если рост происходит на ограниченных во всех направлениях террасах. Рассмотрим формирование массива квантовых проволок (нитей) на примере гетероэпитаксиального роста системы типа GaAs–AlAs, у которой постоянные решеток двух материалов почти совпадают и оба материала неустойчивы относительно фасетирования. Полная энергия системы равна

$$E = E_{\text{пов}} + E_{\text{реб}} + \Delta E_{\text{упр}} + E_{\text{интерфейс}}\,. \tag{2.4}$$

Здесь помимо трех составляющих, которые дают вклад в энергию фасетированной поверхности одного материала, входит энергия границы раздела двух материалов $E_{\text{интерфейс}}$. Как показывает анализ возможных режимов роста, тип растущей структуры определяется тем, смачивает ли осаждаемый материал фасетированную подложку или нет. Возможные варианты структур приведены на рис. 2.9.

Если осаждаемый материал смачивает фасетированную подложку, тогда образуется однородное покрытие (рис. 2.9, *а*). Примером является AlAs, осаждаемый на вицинальную поверхность GaAs(001). Если осаждаемый материал не смачивает подложку, то

изолированные кластеры осаждаемого материала образуются в «канавках» периодически фасетированной поверхности (рис. 2.9, *б*, *в*). Такая ситуация реализуется при осаждении GaAs на вицинальную поверхность AlAs{(001) и (311)} и AlAs на GaAs(311). Эксперименты показали возможность прямого получения изолированных кластеров GaAs на AlAs с последующим заращиванием их слоем AlAs и образованием квантовых нитей.

При большой толщине покрытия «холмы» на поверхности осаждаемого материала образуются над «канавками» подложки. В результате возникает непрерывный слой с модуляцией толщины (рис. 2.9, *г*). На этой основе возможно создание квантовой ямы с модулированной шириной.

Для террас, ограниченных со всех сторон на фасетированной поверхности, возможно получение квантовых точек, поскольку адсорбированные на террасе атомы встраиваются преимущественно в углы террас, примыкающие к вышележащим террасам. Так, например, получены квантовые точки InAs на вицинальной подложке GaAs(001), разориентированной в направлении [010].

2.5.3. Трехмерные массивы когерентно-напряженных островков

Трехмерные массивы когерентно-напряженных островков образуются при выращивании гетероэпитаксиальных структур в режиме роста Странского—Крастанова при не очень большом объеме островка ($\sim 10^6$ атомов). Рост структуры на поверхности подложки начинается с образования тонкого однородного, когерентного с подложкой слоя, который называют смачивающим слоем, т. е. в этом случае структура слоя определяется подложкой и, поскольку постоянные решеток материалов пленки и подложки различаются, пленка находится в напряженном состоянии. Дислокации несоответствия при этом не образуются, что и отмечено термином «когерентный». Смачивающий слой, который образуется в режиме роста Странского—Крастанова, имеет микроскопическую толщину, и его поверхность отличается от поверхности объемного кристалла, состоящего из осаждаемого вещества. Если количество осажденного материала, соответствующее Q монослоям, превышает критическую толщину смачивающего слоя Q_c, то избыточное количество осаждаемого материала ($Q - Q_c$) монослоев формирует островки. В местах образования 3D-островков смачивающий слой исчезает и часть материала, который располагался бы в сплошном слое, переходит в объем островка. Формирование трехмерных островков приводит к уменьшению упругой энергии и энергетически более выгодно, чем образование однородной, сравнительно толстой пленки или островка с дислокациями. Поскольку дислокации несоответствия отсутст-

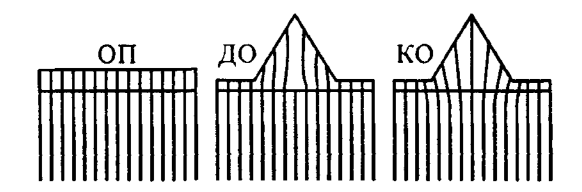


Рис. 2.10. Релаксация упругих напряжений в процессе роста в режиме Странского—Крастанова. Линии обозначают плоскости решетки.

ОП — однородная пленка; ДО — дислокационный островок; КО — когерентный островок.

вуют, то каждая атомная плоскость подложки имеет свое продолжение в объеме островка — это также отмечено термином «когерентный», но из-за различия постоянных решетки материалов подложки и островка возникают механические напряжения и островок становится напряженным. Эксперименты показали, что когерентно-напряженные островки имеют примерно одинаковые размеры с очень малым разбросом по размерам. В системе InAs на подложках из GaAs(001) оптимальный размер островка составляет ~140 Å. Малый разброс по размерам островка наблюдался также в других гетероэпитаксиальных системах: InAs–InP(001), AlInAs–GaAlAs(001), GeSi–Si(001), CdSe–ZnSe(001), InAs–InAlAs(001), InAs–InGaAs (001), InAs–Si(001).

Возможность образования когерентного островка (КО) или островка с дислокациями на границе с подложкой (ДО) определяется отношением энергий боковых поверхностей граней островка $\Delta E_{пов}$ (эта величина всегда положительна) и энергией границы раздела при образовании дислокаций несоответствия $E^{дисл}_{интерфейс}$. На фазовой диаграмме с осями $\Lambda = E^{дисл}_{интерфейс} / \Delta E_{пов}$ и Q — количество монослоев осажденного материала изображены области образования структур разного типа, представленных на рис. 2.10 (рис. 2.11). Как видно из рисунка, если $\Delta E_{пов}$ велика или энергия границы раздела с образованием дислокаций относительно мала, тогда Λ оказывается меньше Λ_0 и образование когерентно-напряженного островка невыгодно. В этом случае при увеличении количества осажденного материала происходит переход от однородной напряженной пленки (ОП) к островку с дислокациями на границе и когерентно-напряженные островки не образуются. Если $\Delta E_{пов}$ мала или энергия границы раздела с образованием дислокаций относительно велика, тогда Λ больше Λ_0 и при увеличении количества осажденного материала происходит переход от однородной пленки к пленке с когерентными островками. Дальнейшее осаждение материала может привести к образованию дислокаций.

Вся система с островками на поверхности, т. е. распределение осажденного на подложку материала Q между смачивающим слоем

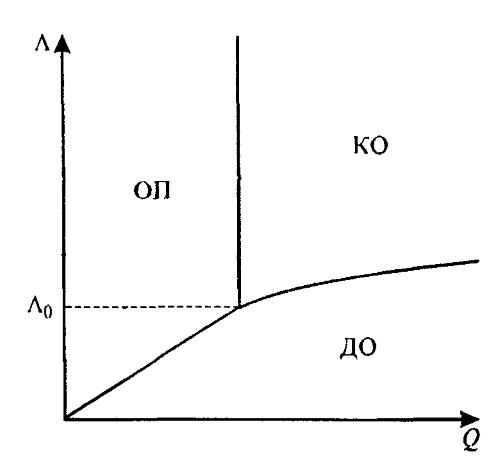


Рис. 2.11. Фазовая диаграмма, показывающая возможные типы структур. Q — количество осажденного материала; Λ — отношение энергий границы раздела с дислокациями и дополнительной поверхностной энергии.

Q_c и островками, форма островка, его объем, распределение островков по поверхности подложки, определяется из условия минимума свободной энергии системы с учетом релаксации упругой энергии при образовании трехмерных островков. Разумеется, сама система, в том числе форма островков, их распределение и т. д., зависит от полного количества осажденного материала. Отличие энергии системы с трехмерными островками от энергии системы со сплошной однородной пленкой на поверхности подложки (при том же количестве осажденного материала) составляет величину

$$\Delta E = W(Q_c) - W(Q) + (Q - Q_c)a\left[-\lambda\varepsilon_0^2 + \frac{\tilde{E}_{остр} - W(Q_c)\tilde{A}_{остр}}{V} + \frac{\tilde{E}_{взаимод}}{2V}\right]. \quad (2.5)$$

Здесь $W(Q)$ — полная энергия системы, состоящей из однородного плоского слоя и подложки, из которой исключена упругая энергия деформации этого слоя; a — постоянная решетки осаждаемого материала; $\varepsilon_0 = \Delta a/a$ — рассогласование решеток двух материалов; λ — упругие модули (для простоты рассматривается кристалл с кубической симметрией); V — объем одного островка; $\tilde{A}_{остр}$ — площадь основания одного островка; $\tilde{E}_{остр}$ — энергия одного отдельного островка; $\tilde{E}_{взаимод}$ — энергия взаимодействия одного островка со всеми остальными островками (зависит от взаимного расположения островков, т. е. от типа сверхрешетки, составленной из островков). Фактически $W(Q)$ — это поверхностная энергия системы подложка—плоский слой. Для сравнительно толстой пленки, когда $Q \gg 1$, $W(Q) = \gamma_2 + \gamma_{12}$, а для тонкой пленки, когда $Q \to 0$, $W(0) = \gamma_1$. Здесь γ — свободная энергия поверхности в расчете на

единицу площади, индекс 1 относится к подложке, а 2 — к слою; γ_{12} — энергия границы раздела подложки и слоя.

Энергию одного островка, которая входит в уравнение (2.5), можно представить в виде суммы трех вкладов:

$$\tilde{E}_{остр} = \tilde{E}_{упр} + \tilde{E}_{пов} + \tilde{E}_{реб}. \qquad (2.6)$$

Энергия упругой деформации островка $\tilde{E}_{упр}$ имеет два источника: с одной стороны, это рассогласование по постоянной решетки между осаждаемым материалом и подложкой, а с другой — скачок тензора поверхностных натяжений на ребрах островка, который приводит к упругой релаксации, аналогичной упругой релаксации при образовании фасетированной поверхности. Взаимодействие этих двух источников тоже дает вклад в упругую энергию островка.

Поверхностная энергия островка $\tilde{E}_{пов}$ зависит от деформации поверхности. Учет этой деформации называется перенормировкой поверхностной энергии. Перенормировка свободной поверхностной энергии учитывает изменение свободной поверхностной энергии из-за деформации поверхности. Перенормировка энергии может привести к уменьшению полной поверхностной энергии, несмотря на увеличение полной площади поверхности при образовании трехмерных (3D) островков.

Последний член (2.6) — короткодействующая энергия ребер островка. Учет всех вкладов в энергию островка приводит к следующему выражению для энергии системы при образовании одного островка:

$$\tilde{E}_{остр}(V) = \Delta\tilde{E}^{V}_{упр} + \Delta\tilde{E}^{ренорм}_{пов} + \Delta\tilde{E}^{реб}_{упр} + \tilde{E}_{реб} =$$
$$= \left[-f_1\lambda\varepsilon_0^2 V + (\Delta\Gamma)V^{2/3} - \frac{f_2\tau^2}{\lambda} V^{1/3} \ln\left(\frac{V^{1/3}}{2\pi a}\right) + f_3\eta V^{1/3} \right]. \qquad (2.7)$$

Здесь первый член в квадратных скобках — энергия объемной упругой релаксации $\Delta\tilde{E}^{V}_{упр}$ и он всегда отрицателен; второй член — изменение перенормированной поверхностной энергии системы из-за образования одного островка $\Delta\tilde{E}^{ренорм}_{пов}$; третий член — вклад ребер островка в энергию упругой релаксации $\Delta\tilde{E}^{реб}_{упр}$, и он всегда отрицателен; четвертый член — короткодействующая энергия ребер $\tilde{E}_{реб}$, где η — характерная энергия на единицу длины ребра; τ — характерная величина тензора собственных поверхностных упругих напряжений; λ — характерная величина тензора модулей упругости; $\Delta\Gamma$ определяется формой островка и зависит от удельной поверхностной энергии боковых граней островка, углов их наклона, энергии границы раздела между островком и подложкой, а также включает в себя энергию взаимодействия двух упругих полей (поля напряжений из-за рассогласования решеток и поля на-

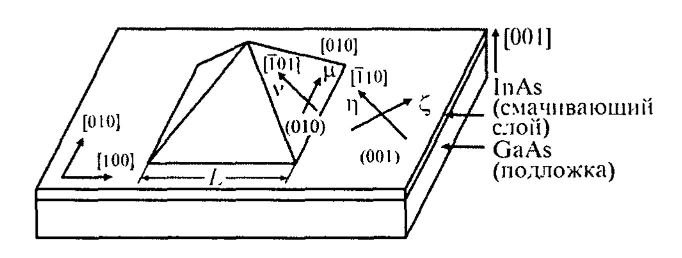


Рис. 2.12. Геометрия пирамиды InAs на поверхности GaAs(001).

пряжений из-за скачка тензора собственных упругих напряжений поверхности), обеспечивающую линейную по величине рассогласования ε_0 перенормировку поверхностной энергии и квадратичную по ε_0 перенормировку; $\Delta\Gamma$ может иметь любой знак; коэффициенты f_1, f_2, f_3 определяются только формой образца и не зависят от его размера.

Равновесное состояние в системе островков достигается за счет обмена веществом между островками путем миграции по поверхности смачивающего слоя. Для разреженной системы островков, когда расстояние между островками велико по сравнению с размером островка, время миграции атомов по поверхности одного островка меньше, чем время миграции между островками. Поэтому равновесная форма одного островка устанавливается быстрее, чем равновесная структура всего массива островков. Для любого объема островка существует равновесная форма и она определяется из условия минимума энергии (2.7) для островков разной формы. При температурах, далеких от температуры плавления, равновесная форма содержит только грани с малой поверхностной энергией. Как правило, это грани с низкими индексами Миллера. В системе InAs–GaAs(001) равновесная форма островка — пирамида с квадратным основанием (рис. 2.12).

В системе разреженных островков упругим взаимодействием между островками можно пренебречь и исключить последний член в квадратных скобках (2.5). В этом случае равновесные размеры островка при условии сохранения его формы можно также найти из условия минимума энергии (2.7). Для случая системы InAs–GaAs (001), когда островок представляет собой пирамиду с квадратным основанием $L \times L$, можно использовать для минимизации энергии уравнение (2.7). Эта энергия зависит от размера островка, углов наклона боковых граней, рассогласования постоянных решетки, энергии ребер, упругих модулей и величины $\Delta\Gamma$. Удобно ввести характерную длину L_0 и характерную энергию E_0, которые зависят от упругих модулей, поверхностных упругих напряжений и от углов наклона граней, а L_0 зависит еще и от энергии ребер, причем $E_0 \propto L_0^{-2}$.

Тогда вклад в энергию островка членов, зависящих от L, запишется в виде $E'(L)$:

$$E'(L) = E_0\left[-2\left(\frac{L_0}{L}\right)\ln\left(\frac{e^{1/2}L}{L_0}\right)+\frac{2\alpha}{e^{1/2}}\left(\frac{L_0}{L}\right)\right], \qquad (2.8)$$

где управляющий параметр α равен отношению изменения в поверхностной энергии из-за образования островка $\Delta E_{\text{пов}}^{\text{ренорм}}$ и энергии вклада ребер в упругую энергию релаксации $\left|\Delta E_{\text{упр}}^{\text{реб}}\right|$:

$$\alpha = \frac{\Delta E_{\text{пов}}^{\text{ренорм}}}{\left|\Delta E_{\text{упр}}^{\text{реб}}\right|} \propto L_0\,(\Delta\Gamma). \qquad (2.9)$$

Для различных значений управляющего параметра α энергия островка в расчете на один атом островка изображена на рис. 2.13. Если $\alpha \le 1$, то существует оптимальный размер островка $L_{\text{опт}}$, соответствующий абсолютному минимуму энергии $E'(L_{\text{опт}}) < 0$. Поскольку объединение всех островков в один большой островок соответствует $L \to \infty$, когда $E'(L) \to 0$, то видно, что в этом случае массив островков оптимального размера $L_{\text{опт}}$ является стабильным и коалесценции островков не происходит. Если $1 < \alpha < 2e^{-1/2} \approx 1.2$, то тогда существует только локальный минимум энергии, соответствующий метастабильному массиву островков при $E'(L) > 0$. Когда $\alpha > 1.2$, локальный минимум энергии $E'(L)$ исчезает. Для всех случаев $\alpha > 1$ существует термодинамическая тенденция к коалесценции. В этом случае минимум энергии соответствует одному гигантскому кластеру, в который собирается осажденный материал.

Если $\Delta\Gamma < 0$ (при этом и $\alpha < 0$), то образование трехмерных островков из-за релаксации приводит не только к уменьшению упругой энергии деформации, но и к уменьшению перенормированной

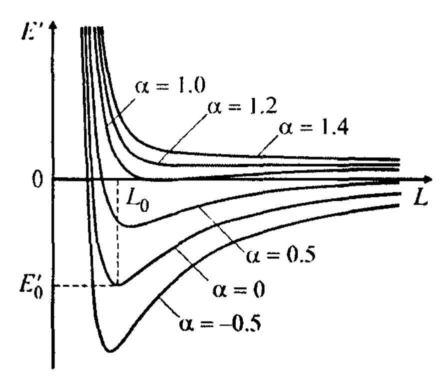


Рис. 2.13. Энергия 3D-когерентно-напряженных островков в расчете на один атом островка в зависимости от размера островка при различных значениях управляющего параметра α.

поверхностной энергии. Изменение поверхностной энергии островка $\Delta\Gamma$ может быть отрицательным даже в том случае, когда плоская поверхность осаждаемого материала сама по себе устойчива относительно спонтанного фасетирования, как например поверхность InAs(001). При возникновении островка появление наклонных граней с площадью большей, чем площадь основания островка, сопровождается исчезновением не участков плоской поверхности осаждаемого материала, а участков смачивающего слоя. Поверхностная же энергия его может быть сильно перенормирована как за счет химических связей с подложкой, так и за счет деформационной зависимости поверхностной энергии. Толщина смачивающего слоя InAs в системе InAs–GaAs(001) имеет микроскопическую толщину в 1—2 монослоя, и ее поверхностная энергия не равна поверхностной энергии InAs(001) в объемном материале.

Для плотного массива островков, когда расстояние между островками сравнимо с размером одного островка, становится существенным их упругое взаимодействие, которое обусловлено проникновением в подложку неоднородного поля напряжений, создаваемого островками. Главным взаимодействием на больших расстояниях в первых порядках приближения является диполь-дипольное взаимодействие

$$\tilde{E}_{\text{взаимод}} \approx F(\varphi)V^2/R^3 , \qquad (2.10)$$

где R — расстояние между двумя островками; $F(\varphi)$ — зависящий от азимутального угла φ в плоскости подложки множитель, определяемый упругой анизотропией подложки. Результирующее взаимодействие между островками на поверхности является отталкиванием при любом направлении между островками,[5] что обеспечивает устойчивость массива островков. Вид решетки, в которой упорядочены островки массива, зависит от доли поверхности, занятой островками. Расчет упругой энергии для нескольких латеральных типов сверхрешеток приводит к выводу, что наиболее выгодной на поверхности (001) упругоанизотропного кубического кристалла[6] является двумерная квадратная сверхрешетка (когда доля покрытия поверхности равна или превышает 0.33). Векторы

[5] В системах с сильной упругой анизотропией диполь-дипольное взаимодействие меняет знак в зависимости от ориентации отрезка прямой, соединяющего два островка, относительно кристаллографических осей подложки. Когда направление отрезка близко к направлениям наилегчайшего сжатия [100] или [010], отталкивание сменяется притяжением. Поэтому разреженный массив островков выстраивается в слабо связанные цепи, параллельные направлениям [100] или [010]. При этом учет более высоких порядков приближения приводит к результирующему отталкиванию.

[6] Большинство материалов $A^{III}B^{V}$, $A^{II}B^{VI}$, Si и Ge обладают сильной анизотропией упругости.

примитивных трансляций решетки направлены по осям наилегчайшего сжатия [100] и [010]. Энергетическая выгодность квадратной решетки определяется именно упругой анизотропией подложки, а не формой отдельного островка. Для упругоизотропной среды более предпочтительной является гексагональная сверхрешетка.

Таким образом, возможно существование равновесного массива островков, имеющих оптимальный размер и упорядоченных в квадратную решетку.

Наиболее изученной в экспериментальном отношении является система трехмерных островков InAs–GaAs(001). Когерентно-напряженные островки InAs на подложке GaAs получены методами МЛЭ и газофазной эпитаксии из МОС. Исследование морфологии системы, полученной методом МЛЭ, показало следующее. Когда средняя толщина осажденного InAs достигает 1.6—1.7 монослоя (МС), то происходит образование трехмерных островков InAs. Эта толщина является критической для начала образования трехмерных островков. Осаждение 2 МС InAs приводит к образованию островков малого размера, в основном не проявляющих четкой кристаллографической огранки и обладающих большим разбросом по размерам. Увеличение толщины осажденного InAs до 4 МС приводит к образованию плотного массива островков, имеющих квадратное основание со сторонами, ориентированными по осям [100] и [010]. Средние размеры точек составляют ~140 Å и являются равновесными размерами островка. Высота островка составляет ~60 Å. Исследования методом просвечивающей электронной спектроскопии высокого разрешения показало, что островки InAs имеют форму пирамиды с наклонными гранями (010) и острыми ребрами между гранями. Статистическое исследование взаимного расположения островков показало преимущественное направление между ближайшими соседями вдоль осей [100] и [010].

Аналогичным образом растут островки в системе $Si_{1-x}Ge_x$–Si(100) [21]. В этом случае толщина смачивающего слоя составляет 4—5 МС германия. Затем, при увеличении количества осажденного германия, образуются ограненные пирамидки (hut-островки), которые при дальнейшем росте превращаются в куполообразные островки (dome-островки). Dome-островки достигают предельного размера при осаждении германия в количестве 11 МС. Доля кремния в островках доходит до 50 % в результате диффузии Si из подложки в процессе роста. Средняя предельная высота составляет 20 нм. Поверхностная плотность островков порядка $3 \cdot 10^9$ см$^{-2}$.

Процесс образования островков InAs в значительной мере зависит от давления мышьяка в камере роста. Изменяя давление мышьяка от стандартного для МЛЭ значения ($p^0_{As} \approx 2 \cdot 10^{-6}$ Торр) в сторону уменьшения до значений (1/6) p^0_{As}, можно получить вместо трехмерных островков плоские островки InAs с размерами ~1000 Å.

При росте давления до $3p^0_{As}$ концентрация точек уменьшается и появляется высокая концентрация больших (~500—1000 Å) кластеров InAs, содержащих дислокации. Все изменения морфологии системы в зависимости от давления мышьяка объясняются на основе поверхностных энергий. Изменение давления As_2 вызывает изменение поверхностной энергии смачивающего слоя и приводит к режиму роста, когда имеется тенденция к коалесценции. Это происходит при повышении давления мышьяка. При низком давлении мышьяка In образует на поверхности GaAs сегрегационный слой, который можно рассматривать как квазижидкую фазу, поверхностная энергия которой может сильно отличаться от поверхностной энергии кристаллического InAs.Такое изменение поверхностной энергии меняет критическую толщину напряженного слоя, при которой происходит образование трехмерных островков. Экспериментально установлено, что InAs растет на GaAs(001) в условиях избытка In псевдоморфно до толщины 7 МС, что и вызывает образование плоских островков толщиной несколько МС.

Для целей практического использования в лазерных структурах изготавливают многослойные массивы квантовых точек. Для этого слой трехмерных когерентно-напряженных островков, состоящих из материала 2 на подложке из материала 1, заращивают сверху слоем материала 1. Заращенные островки представляют собой когерентные упругие включения, которые создают дальнодействующие поля упругих напряжений во всей заращенной гетерофазной системе. При повторном осаждении материала 2 на заращенную гетероструктуру возникает принципиально новый режим роста: рост в поле упругих напряжений, созданных заращенными точками первого слоя. Поскольку диффузией в объеме можно пренебречь по сравнению с поверхностной, так как коэффициенты объемной диффузии атомов основных компонентов в материалах $A^{III}B^{V}$ при характерных температурах роста на несколько порядков меньше коэффициентов поверхностной диффузии, то заращенные островки создают статическое поле упругих напряжений. Упругая деформация возрастает с ростом объема островка и рассогласования постоянных решеток между материалами островка и матрицы. Какой характер имеет деформация — растяжение или сжатие, зависит от того, какой параметр решетки больше: у материала островка или матрицы. Если материал островка имеет больший параметр решетки, чем у матрицы (как в случае систем InAs–GaAs и Ge–Si), то деформация поверхностного слоя матрицы имеет характер растяжения. В этом поле упругих напряжений и происходит поверхностная миграция адсорбированных атомов. Модулированное поле деформаций на поверхности приводит к модуляции химического потенциала поверхностных адсорбированных атомов, что в свою очередь приводит к их миграции в виде

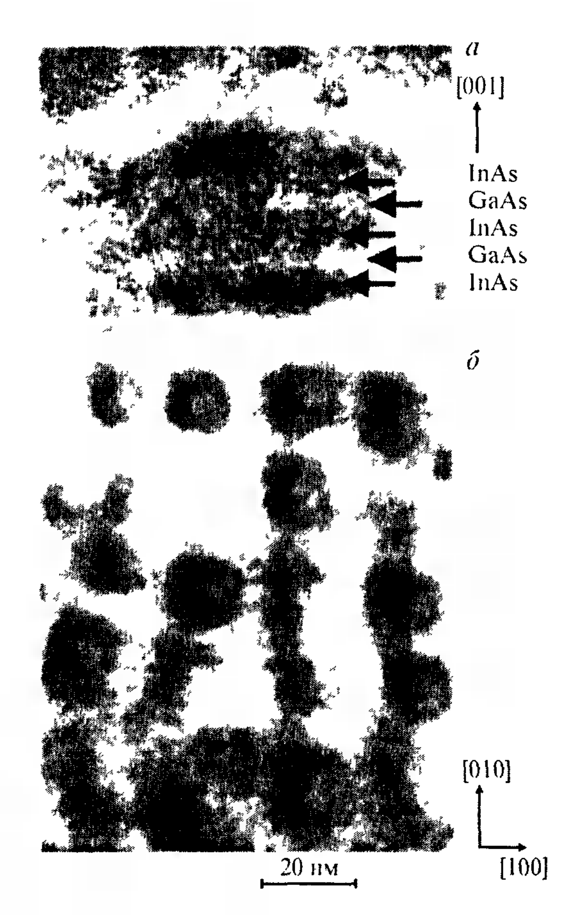


Рис. 2.14. Вертикально связанные квантовые точки InAs в матрице GaAs. Изображение получено методом просвечивающей электронной микроскопии высокого разрешения в поперечном сечении вдоль направления [010] массива туннельно-связанных в z-направлении ([100]) квантовых точек.

а — вид в сечении, *б* — вид сверху на плоскость подложки.

диффузии и дрейфа. Этот дрейф является движущей силой кинетической самоорганизации в таком сложном режиме роста.

Для островка InAs в матрице GaAs положение на поверхности непосредственно над островком соответствует максимальному растяжению. Поскольку атомы In имеют больший радиус, чем атомы Ga, то In притягивается к области поверхности с максимальной деформацией растяжения, т. е. к положениям на поверхности матрицы прямо над заращенным островком. Экспериментальные данные и теоретические модели показывают, что латеральное упорядочение (вдоль поверхности) в последующих слоях проявляется даже лучше, чем упорядочение первого слоя.

Многослойные массивы трехмерных когерентно-напряженных островков образуют новый класс спонтанно упорядоченных наноструктур, в которых упорядочение происходит как в направлениях плоскости подложки, так и в вертикальном направлении.

Для применения в оптоэлектронике и управления электронными свойствами массивов квантовых точек изготавливают массивы вертикально связанных квантовых точек. С этой целью заращивающий слой GaAs осаждают таким образом, чтобы пирамида InAs была закрыта только частично, и затем снова осаждают InAs. Так производят несколько циклов осаждения InAs–GaAs. Тогда точки InAs, вырастающие над точками нижнего слоя, оказываются электронно-связанными с точками нижнего слоя. На рис. 2.14 изображены вертикально связанные квантовые точки InAs в матрице GaAs. Размеры островков в вышележащих слоях возрастают и достигают ~170 Å, в то время как в первом слое размер островка составляет ~110 Å.

2.5.4. Поверхностные структуры плоских упругих доменов

Существуют различные типы плоских образований на поверхности кристалла или на поверхности подложки в гетерофазной структуре, у которых размеры вдоль плоскости поверхности значительно превосходят их высоту. Такие структуры возникают, например, когда на поверхности могут сосуществовать различные фазы поверхностной реконструкции или на поверхности подложки в гетероэпитаксиальных системах при образовании островков монослойной высоты в случае субмонослойного покрытия, о которых уже упоминалось в предыдущем разделе в случае системы InAs–GaAs(001).

Поскольку атомы на поверхности кристалла находятся в иных условиях, чем в объеме, то в некоторых кристаллах расположение атомов на поверхности отличается от объемного, что обеспечивает минимизацию свободной энергии. Отличие структуры атомов на поверхности от объемного их расположения называют реконструкцией поверхности и структуру атомов на поверхности обозначают как $m \times n$. Реконструкция $m \times n$ означает, что два основных вектора на поверхности в m и n раз соответственно больше, чем аналогичные векторы в объеме. Элементарная ячейка поверхностного слоя в $m \times n$ раз больше по сравнению с эквивалентной элементарной ячейкой слоя, расположенного в объеме. На поверхности Si(001), например, могут сосуществовать фазы поверхностной реконструкции 2×1 и 1×2, причем они разделены ступенями моноатомной высоты.

Как в случае разных фаз реконструкции на поверхности, так и в случае моноатомных островков эти образования представляют со-

бой домены, которые образуют стабильные периодические структуры на поверхности. Домены и окружающие их области различаются между собой по механическим свойствам и, в частности, обладают разными значениями тензора собственных поверхностных напряжений $\tau_{\alpha\beta}$. На доменных границах возникают эффективные силы, приложенные к кристаллу. Эти силы создают поле упругих деформаций и дают вклад в релаксацию упругой энергии. Полная энергия системы плоских доменов в расчете на единицу площади равна:

$$E = E_{\text{пов}} + E_{\text{гран}} + \Delta E_{\text{упр}}. \qquad (2.11)$$

Поверхностная энергия $E_{\text{пов}}$ не зависит от периода структуры D, энергия доменных границ равна $E_{\text{гран}} = C_1\eta D^{-1}$, где η — короткодействующая энергия доменных границ, и упругая энергия релаксации имеет вид $\Delta E_{\text{упр}} = -C_2(\Delta\tau)^2 Y^{-1} D^{-1} \ln(D/a)$, где Y — модуль Юнга, C_1 и C_2 — геометрические факторы. Благодаря логарифмической зависимости энергии упругой релаксации от периода структуры полная энергия всегда имеет минимум при определенном оптимальном периоде:

$$D_{\text{опт}} = a \exp\left[\frac{C_1 \eta Y}{C_2(\Delta\tau)^2} + 1\right]. \qquad (2.12)$$

В случае сосуществования на поверхности фаз поверхностной реконструкции образуются длинные полоски с периодом следования по одному из направлений на поверхности $D_{\text{опт}}$ (рис. 2.15). Такая структура может быть использована для изготовления сверхрешеток квантовых нитей в случае осаждения материала с меньшей шириной запрещенной зоны, чем у подложки с плоскими доменами.

Важным типом планарных структур поверхностных доменов являются массивы периодически расположенных островков, образованные в гетероэпитаксиальных системах при субмонослойном покрытии. Периодические структуры узких длинных полосок наблюдались в системе O–Cu(110). Для полупроводниковой наноэлектроники более важной системой при субмонослойном покрытии является система InAs–GaAs(001). В этом случае на поверхно-

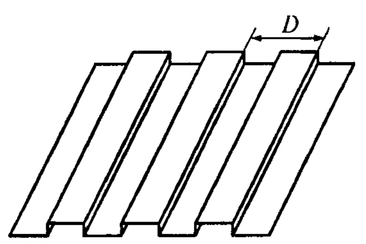


Рис. 2.15. Периодическая структура плоских доменов.

сти образуются монослойные островки InAs в виде узкой, длинной полосы, вытянутые в направлении [1$\bar{1}$0]. Ширина островка составляет 4 нм. Эта система островков является термодинамически равновесной и устойчивой.

Поверхностные структуры планарных доменов появляются сами собой, когда покрытие достигает или превосходит 1 МС. Так, в системе InAs–GaAs(001) сплошное монослойное покрытие InAs неустойчиво и распадается на участки поверхности, покрытые островками высотой 1 или 2 МС, и области чистой поверхности подложки GaAs. Рельеф поверхности повторяется в направлении [100] и хорошо различим для поверхностного покрытия 1—1.5 МС до того, как происходит переход к росту трехмерных островков, который начинается при среднем покрытии ~1.7 МС.

Симметрия равновесных структур в виде островков монослойной высоты определяется как симметрией тензора собственных поверхностных напряжений $\tau_{\alpha\beta}$, так и симметрией объемных упругих модулей подложки. Главные оси тензоров собственных поверхностных напряжений GaAs и InAs — [1$\bar{1}$0] и [110], объемные упругие свойства материалов определяются направлениями осей наилегчайшего сжатия [100] и [010]. По мере роста количества осажденного InAs происходит изменение ориентации доменов от направления [1$\bar{1}$0], характерного для субмонослойного покрытия, к ориентации по направлениям [100] и [010] при покрытии 1.0—1.5 МС. Этот переход является следствием возрастания вклада рассогласования по постоянной решетки в поле упругих напряжений по сравнению с вкладом, обусловленным скачком $\tau_{\alpha\beta}$ на границе двух фаз. Это возрастание связано с увеличением объема осажденного InAs.

Для оптоэлектронных применений необходимы многослойные массивы островков монослойной высоты. Их получают так же, как и в случае трехмерных островков, путем заращивания островков InAs слоем GaAs. Можно было бы ожидать такого же расположения островков, как и в трехмерном случае. Однако эксперименты показали, что островки следующего слоя растут не коррелированно над островками нижнего слоя, а смещены относительно их, и даже возможна антикорреляция в расположении островков. Такая особенность роста объясняется тем, что полная энергия поверхностного массива островков в поле упругой деформации заращенных островков равна:

$$E = E_{\text{пов}} + E_{\text{гран}} + \Delta E_{\text{упр}}^{\text{пов}} + E_{\text{упр}}^{\text{об}}, \qquad (2.13)$$

где $E_{\text{пов}}$ — сумма поверхностной энергии поверхности островка и непокрытой части материала матрицы; $E_{\text{гран}}$ — энергия границ островка; $\Delta E_{\text{упр}}^{\text{пов}}$ — энергия упругой релаксации поверхности островков из-за скачка тензора собственных упругих напряжений на границах островка; $E_{\text{упр}}^{\text{об}}$ — упругая энергия взаимодействия между

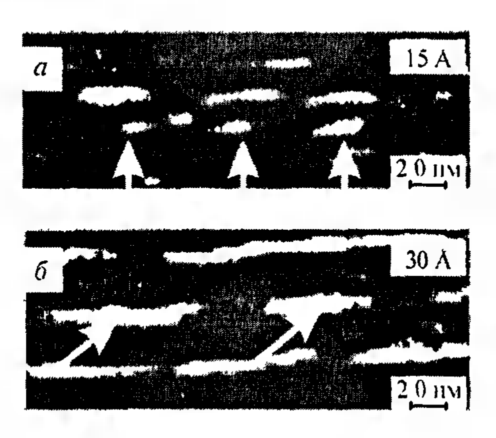


Рис. 2.16. Изображение многослойного массива двумерных островков CdSe, погруженных в матрицу ZnSe.

а — последовательные слои разделены слоем материала матрицы толщиной 15 Å (наблюдается вертикальная корреляция островков), *б* — последовательные слои разделены слоем материала матрицы толщиной 30 Å (наблюдается состояние антикорреляции)

поверхностными и заращенными островками. Первые три члена уравнения (2.13) определяют периодическую структуру массива островков, которая соответствует минимальному значению суммы этих членов. Если бы взаимодействием поверхностных и заращенных островков можно было пренебречь, то поверхностный массив островков как целое был бы смещен произвольным образом относительно островков нижнего слоя. Поле деформаций заращенных островков имеет ту же самую периодичность, что и массив поверхностных островков, поэтому четвертый член (2.13) не меняет периодичность поверхностной структуры и как раз определяет ее положение по отношению к массиву заращенных островков. Расчет показывает, что энергия взаимодействия $E_{\text{упр}}^{\text{об}}$ убывает по мере увеличения толщины заращивающего слоя материала матрицы. Кроме того, упругие деформации, вызванные заращенными островками, также убывают по мере увеличения толщины заращивающего слоя и при этом осциллируют. Как следствие взаимодействие между последовательными слоями островков имеет вид осциллирующего спада по мере увеличения расстояния между слоями. В результате в зависимости от толщины заращивающего слоя матрицы можно получить и коррелированное состояние, когда островки следующего слоя располагаются непосредственно над островками нижнего слоя, и состояние антикорреляции (островки верхнего слоя располагаются в промежутках между островками нижнего слоя), и промежуточное состояние. На рис. 2.16 изображено сечение структуры CdSe–ZnSe, в которой в зависимости от толщины разделяющего слоя наблюдается состояние корреляции (рис. 2.16, *а*)

и антикорреляции (рис. 2.16, *б*). Аналогичные структуры получены в системе InAs– GaAlAs. Такого рода субмонослойные сверхрешетки имеют хорошие перспективы для применения в оптоэлектронике.

2.5.5. Структуры с периодической модуляцией состава в эпитаксиальных пленках твердых растворов полупроводников

Еще одна возможность получения наноструктур — это использование неустойчивости твердых растворов относительно спинодального распада [22]. Неустойчивость твердого раствора $A_{1-c}B_cC$ означает, что твердый раствор с некоторым неоднородным профилем состава $c(\mathbf{r}) = \bar{c}(\mathbf{r}) + \delta c(\mathbf{r})$ имеет меньшую свободную энергию, чем свободная энергия однородного твердого раствора $F = \int[H(\bar{c}) - TS_{см}(\bar{c})]dV$ с составом $c(\mathbf{r}) = \bar{c}(\mathbf{r})$, где $S_{см}$ — энтропия смешивания, T — температура. Неустойчивость однородного твердого раствора возникает, когда изменение энтальпии при образовании твердого раствора $A_{1-c}B_cC$ $\Delta H_{обр}$ из бинарных компонентов AC и BC положительно, т. е. $\Delta H_{обр} = H(A_{1-c}B_cC) - (1-c)H(AC) - cH(BC) > 0$.[7] Тогда при $T = 0$ двухфазная смесь чистых материалов AC и BC имеет меньшую свободную энергию, чем однородный твердый раствор $A_{1-c}B_cC$, и он оказывается неустойчивым. При конечных значениях температуры T вклад энтропии в свободную энергию способствует перемешиванию компонентов и стабилизирует однородный твердый раствор.

Абсолютный минимум энергии имеет совершенный кристалл с абсолютной периодичностью, а твердые растворы — системы неупорядоченные. Атомы в твердом растворе занимают кристаллические узлы, но атомы разного типа расположены хаотично. Поэтому существуют скопления атомов одного типа, которые соответствуют локальному изменению концентрации $c = \bar{c} + \delta c(r)$. В соответствии с правилом Вегарда постоянная решетки зависит от локального состава твердого раствора: $a(c) = a(\bar{c}) + (\partial a/\partial c)\big|_{c=\bar{c}}\,\delta c(\mathbf{r})$, т. е. области твердого раствора с различным составом имеют различные значения постоянной решетки. На границе областей с разным составом возникает упругая деформация, с которой связана упругая энергия $E_{упр}$. При образовании флуктуации состава $\delta c(\mathbf{r})$ изменение свободной энергии равно:

$$\delta F = \int\{[H(\bar{c} + \delta c(\mathbf{r})) - TS_{см}(\bar{c} + \delta c(\mathbf{r}))] - [H(\bar{c}) - TS_{см}(\bar{c})]\}\,dV + E_{упр}. \quad (2.14)$$

---

[7] Для всех тройных твердых растворов полупроводников $A^{III}B^{V}$ $\Delta H_{обр} > 0$

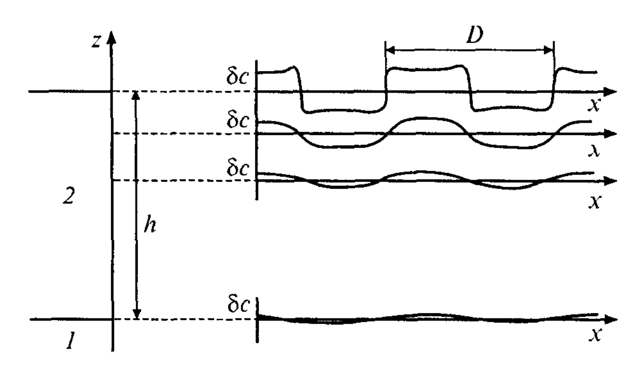


Рис. 2.17. Равновесный профиль состава твердого раствора в эпитаксиальной пленке (*2*) растущей на подложке (*1*).

Поскольку $E_{упр} = 0$ в однородном твердом растворе и $E_{упр} > 0$ в неоднородном твердом растворе, то упругая энергия стабилизирует однородный твердый раствор. Расчет показывает, что объемные тройные твердые растворы полупроводников $A^{III}B^{V}$ устойчивы относительно спинодального распада при всех температурах.

В эпитаксиальных пленках твердых растворов, согласованных по постоянной решетки с подложкой, ситуация меняется. Вблизи свободной поверхности происходит релаксация напряжений, что уменьшает эффект упругой стабилизации однородного твердого раствора. На поверхности ($z = 0$) возникает волна флуктуации состава, которая затухает экспоненциально в глубину пленки: $\delta c(\mathbf{r}) \sim \exp(-|kz|)\exp(i\mathbf{k}_{\parallel}\mathbf{r}_{\parallel})$. Волновой вектор в плоскости поверхности кубического кристалла (001) $\mathbf{k}_{\parallel} \equiv (k_x, k_y)$ направлен вдоль оси наилегчайшего сжатия [100] или [010], поскольку именно в этих направлениях из-за анизотропии модуль Юнга имеет минимальные значения (и максимален коэффициент растяжения), а вектор $\mathbf{r}_{\parallel}$ лежит в плоскости поверхности. Релаксация упругих напряжений вблизи свободной поверхности уменьшает упругую энергию на множитель ~1/3 по сравнению с упругой энергией в объемном образце. В результате многие твердые растворы оказываются неустойчивыми относительно спинодального распада уже

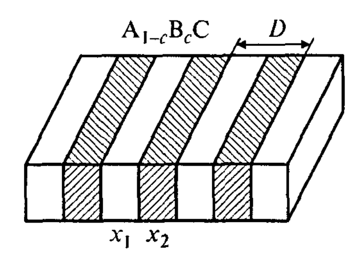


Рис. 2.18. Структура с модуляцией состава твердого раствора.

при 450—500 °C. Возникает равновесная структура с модуляцией состава, для которой состав модулирован вдоль направления [100] или [010], а амплитуда модуляции максимальна на свободной поверхности и затухает в глубь пленки (рис. 2.17), период модуляции D сравним с толщиной эпитаксиальной пленки h: $D \cong h$. Таким образом, релаксация упругих напряжений вблизи свободной поверхности способствует спонтанному образованию наностуктур, а упругая анизотропия материала определяет ориентацию наноструктур.

Поскольку модуляция состава происходит только по одному направлению на поверхности, то на поверхности образуются длинные нити с постоянным составом (рис. 2.18). Состав материала определяет ширину его запрещенной зоны, и, следовательно, в приповерхностном слое возникает модуляция ширины запрещенной зоны по направлению [100] или [010]. В области, где состав материала имеет меньшую ширину запрещенной зоны, на поверхности эпитаксиальной пленки образуются квантовые нити.

2.6. Нанотехнологии

Описанные в предыдущих разделах методы получения наноструктур в той или иной мере используют элементы хорошо развитой традиционной технологии. В то же время развитие техники привело к появлению оборудования, которое позволяет непосредственно изготовить наноструктуры [23]. Эти возможности появились после изобретения сканирующего туннельного микроскопа (СТМ) и атомно-силового микроскопа (АСМ) [24, 25].

С помощью иглы СТМ возможно перемещать отдельные атомы и располагать их в нужной последовательности. Если атом, который необходимо переместить, соединен с поверхностью образца ковалентной связью, то электрического поля, создаваемого иглой СТМ, может оказаться достаточно для разрыва этой связи. В результате такого акта атомной эмиссии атом будет прицеплен к кончику иглы. Затем иглу СТМ перемещают в другую точку поверхности, меняют знак напряжения смещения между иглой и поверхностью образца и помещают атом обратно на поверхность. Так, например, группой исследователей из IBM на поверхности кристалла никеля была изготовлена из отдельных атомов ксенона надпись «IBM». Однако любая созданная таким образом наноструктура может легко разрушаться из-за миграции атомов по поверхности. Именно по этой причине упомянутая выше надпись «IBM» была изготовлена только при гелиевой температуре. Тем не менее работы в этом направлении нанотехнологии продолжаются: в работе [26] сообщалось о перемещении и закреплении на подложке молекул при комнатной температуре.

Другой способ использования СТМ в нанотехнологии состоит в том, что между зондом и подложкой прикладывают достаточно большое напряжение, под действием которого материал с кончика зонда начинает напыляться на подложку. При перемещении зонда атомы попадают на поверхность и образуют нить нанометрового масштаба. Предложен также способ разложения металлоорганических соединений при прохождении тока между иглой СТМ и подложкой. В результате металл осаждается на подложку, и таким образом при движении иглы тоже образуют тонкую металлическую нить субмикронного размера. В результате становится возможным изготавливать из нанопроволок рисунки сложной формы, создавать элементы наноструктур. Для одновременного изготовления большого количества наноструктур на одной подложке разрабатывают наноманипуляторы, содержащие большое число микро-СТМ. Управление наноманипуляторами производится с помощью компьютера, они способны одновременно с изготовлением прибора вести наблюдение за поверхностью, на которой производится сборка наноструктур.

Применение зонда АСМ позволяет использовать непроводящие подложки, он может перемещать атомы чисто механически. Это позволяет разработать новый тип нанолитографии — с применением АСМ [27, 28]. Таким способом был изготовлен МОП-транзистор с общим размером 100 нм и с рабочей областью 60 нм.

Изготовление наноструктур и наноприборов с помощью СТМ и АСМ можно назвать механическим синтезом (механосинтезом). Эти работы находятся в самой начальной стадии развития, и потребуется еще много усилий для доведения методов механосинтеза до уровня промышленного производства.

Другое направление развития в изготовлении компонентов для электроники нанометрового масштаба — это химический синтез (химическая самосборка) [29], в котором используются супермолекулярные структуры. Химическая самосборка — это спонтанная ориентация ряда молекул в энергетически выгодную супермолекулярную структуру. На этой основе возможно изготовление элементов молекулярной электроники, в которых, например, выполнение операций переключения производится за счет перемещения атома. Были выполнены успешные опыты по использованию биологических молекул для изготовления наноразмерных объектов. В них использовали малые органические молекулы с функциональной группой на одном конце, с помощью которой они прикрепляются к подложке в один ряд (так называемые самособранные монослои (SAM — Self-Assembled Monolayers)). Они могут быть использованы вместо фоторезиста для защиты поверхности от травящих растворов. В SAM удается получить отверстия нанометрового масштаба и изготовить рисунок, который невозможно создать при по-

мощи обычной литографии. Известны эксперименты по изготовлению упорядоченных нанокластеров CdS с использованием пептидов. Эксперименты в направлении химического синтеза также находятся в самой начальной стадии развития. Интерес к этому виду нанотехнологии обусловлен тем, что химическая самосборка дает возможность параллельной сборки чрезвычайно большого количества совершенно одинаковых элементов. Для нее также характерен процесс исправления ошибок: в процессе сборки система сама имеет тенденцию освобождать растущую структуру от каких-либо молекул «ошибочного типа» или в «ошибочном положении».

Глава 3

СВОБОДНЫЕ И СВЯЗАННЫЕ НОСИТЕЛИ В НАНОСТРУКТУРАХ

В главе 1 мы обсудили важнейшую характеристику носителей в квантово-размерных структурах — их энергетический спектр. Знание спектра позволяет рассчитать все равновесные электронные свойства системы, что и будет сделано в данной главе. Мы получим выражения для плотности состояний в системах различной размерности и изучим распределение равновесных электронов и дырок внутри квантово-размерных подзон и между ними. В последнем разделе главы будет также рассмотрен вопрос о видоизменении указанных спектров под влиянием кулоновского взаимодействия электронов с примесными центрами и дырками, приводящего к возникновению связанных примесных состояний и экситонов.

3.1. Плотность состояний в электронных системах пониженной размерности

Важной характеристикой любой электронной системы наряду с ее энергетическим спектром является плотность состояний $g(E)$. Напомним, что в массивном образце с параболическим законом дисперсии $g(E)$ возрастает с увеличением энергии от края разрешенной зоны как $\sqrt{E}$ (рис. 3.1, *а*). Мы начнем данный раздел с нахождения $g(E)$ для двумерной электронной системы с энергетическим спектром (1.4) [1]. Для этого предварительно вычислим вспомогательную функцию $G(E)$ — полное число состояний в интервале энергий от 0 до E.

Очевидно, что минимальная энергия, которую могут иметь электроны, равна E_1 и потому $G(E) = 0$ при $E < E_1$. Рассмотрим область энергий $E_1 < E < E_2$. Такие энергии могут иметь только электроны первого уровня, чей полный импульс $p = (p_x^2 + p_y^2)^{1/2} < [2m(E - E_1)]^{1/2}$. Эти электроны занимают в четырехмерном фазовом пространстве $(x,\ y,\ p_x,\ p_y)$ объем, равный $L_x L_y \pi p^2 = 2\pi L_x L_y m(E - E_1)$, где L_x и L_y — размеры образца в плоскости двумерного электронного газа. Согласно принципам статистической физики, на каждое состояние двумерного свободного движения в

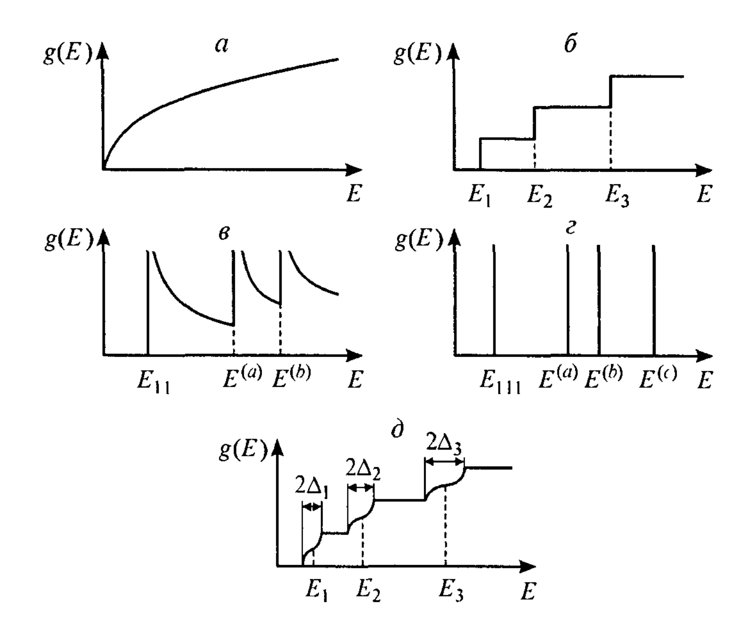


Рис. 3.1. Зависимость плотности электронных состоний от энергии в массивных полупроводниках (*а*), квантовых ямах (*б*), квантовых нитях (*в*), квантовых точках (*г*) и сверхрешетках (*д*) ($E^{(a)}$, $E^{(b)}$, $E^{(c)}$ — энергетические уровни размерного квантования в квантовых нитях и точках, лежащие выше уровней основного состояния E_{11} и E_{111}).

указанном фазовом пространстве приходится элементарный объем $(2\pi\hbar)^2$. Поэтому полное число состояний с энергией, меньшей E, с учетом двукратного спинового вырождения дается формулой

$$G(E) = \frac{mL_xL_y}{\pi\hbar^2}(E - E_1). \tag{3.1}$$

Интересующая нас величина $g(E)$, определяющая число состояний в единичном интервале вблизи заданной энергии, определяется производной от G по энергии. Кроме того, $g(E)$ определяют как число состояний не в образце в целом, а в расчете на единицу площади двумерного электронного газа. Поэтому окончательно

$$g(E) = \frac{1}{L_xL_y}\frac{dG}{dE} = \frac{m}{\pi\hbar^2} \quad (E_1 < E < E_2). \tag{3.2}$$

При энергиях, больших E_2, возможно существование электронов не только в первой, но и в вышележащих подзонах. Каждая подзона будет давать дополнительный вклад в $g(E)$, такой же, как (3.2). Поэтому плотность состояний будет испытывать скачки, равные $m/\pi\hbar^2$, каждый раз, когда энергия электронов сравнивается

с дном очередной подзоны E_n. Это позволяет обобщить (3.2) на случай произвольной энергии E:

$$g(E) = \frac{m}{\pi\hbar^2}\sum_n \Theta(E - E_n), \tag{3.3}$$

где $\Theta(x)$ — единичная функция Хевисайда, равная единице при $x > 0$ и нулю при $x < 0$. На рис. 3.1, *б* приведен график этой ступенчатой функции.

Аналогично можно найти плотность состояний для квантовых нитей с энергетическим спектром (1.5) [6]. В интервале энергий между низшим E_{11} и следующим по энергии квантовым уровнем энергиям, меньшим, чем некоторое заданное E, отвечает область импульсов $-[2m\,(E - E_{11})]^{1/2} < p_x < [2m\,(E - E_{11})]^{1/2}$. Соответствующий объем фазового пространства (x, p_x) равен $2L_x\,[2m(E - E_{11})]^{1/2}$, а элементарный объем на одно состояние составляет $2\pi\hbar$. С учетом спинового вырождения это дает

$$G(E) = \frac{2L_x\sqrt{2m(E - E_{11})}}{\pi\hbar}. \tag{3.4}$$

Вычисляя производную и суммируя вклады всех подзон, получаем окончательно плотность состояний в расчете на единицу длины нити:

$$g(E) = \frac{\sqrt{2m}}{\pi\hbar}\sum_{m,n}\frac{\Theta(E - E_{mn})}{\sqrt{E - E_{mn}}}. \tag{3.5}$$

Эта функция, изображенная на рис. 3.1, *в*, имеет особенности (обращается в бесконечность) при энергиях, соответствующих квантовым уровням в нити.

Наконец, в случае квантовых точек, где энергетический спектр носит чисто дискретный характер, $G(E)$ увеличивается на единицу каждый раз, когда E сравнивается с каким-либо квантовым уровнем.[1] При этом $g(E)$ описывается с помощью так называемой дельта-функции Дирака $\delta(x)$, которая равна нулю при $x \neq 0$, а при $x = 0$ обращается в бесконечность так, что интеграл $\int \delta(x)\,dx = 1$ при любом интервале интегрирования, содержащем точку $x = 0$. Плотность состояний

$$g(E) = \sum_{l,m,n}\delta(E - E_{lmn}) \tag{3.6}$$

---

[1] Если, как это часто бывает в точках, имеющих симметричную форму, уровни энергии являются вырожденными, то величина скачка G равна кратности вырождения данного уровня.

представляет собой совокупность бесконечно узких и высоких пиков, схематично показанных на рис. 3.1, *г*.

Рис. 3.1 представляет идеализированную картину. На самом деле наличие рассеяния носителей приводит к уширению энергетических уровней (см. раздел 1.2). В результате ступеньки на рис. 3.1, *б* приобретут конечную ширину скачка $\sim\hbar/\tau$, а бесконечные пики на рис. 3.1, *в*, *г* станут конечными по высоте и ширине с высотой, тем меньшей, чем ниже подвижность носителей.

Следует отметить, что плотность состояний в формулах (3.3), (3.5) и (3.6) имеет различную размерность. Хорошо известно, что в массивных трехмерных полупроводниках характеристикой, определяющей физические свойства вещества вне зависимости от объема образца V, является не полное число электронов в образце N, а их объемная концентрация N/V. Соответственно плотность состояний g также определяет объемную концентрацию и имеет размерность (эрг · см^3)$^{-1}$ или (Дж · м^3)$^{-1}$ в зависимости от выбранной системы единиц. Двумерные структуры являются однородными по площади, но не по объему. Поэтому полное число электронов N пропорционально не объему, а площади образца S и все физические свойства определяются концентрацией на единицу площади $n_S = N/S$. В результате реальный смысл имеет плотность состояний в расчете на единицу площади (3.3) с размерностью (эрг · см^2)$^{-1}$ или (Дж · м^2)$^{-1}$. Аналогично плотность состояний в квантовых нитях (3.5) вычисляется в расчете на единицу длины и имеет размерность (эрг · см)$^{-1}$ или (Дж · м)$^{-1}$. Наконец, в квантовых точках все определяется абсолютным числом электронов в точке и плотность состояний (3.6) имеет размерность просто обратной энергии.

Можно вычислить и плотность состояний для свехрешеток со спектром [10]. Не приводя здесь непосредственного вывода, дадим лишь конечный результат. В областях энергий, соответствующих запрещенным минизонам, $g(E) = \mathrm{const}\,(E)$, как в чисто двумерном случае (3.3), а внутри разрешенных минизон плотность состояний плавно меняется, имея бесконечную производную на краю минизоны:

$$g(E) = \frac{m}{\pi\hbar^2}\left(n - 1/2 + \frac{1}{\pi}\arcsin\frac{E - E_n}{\Delta_n}\right), \quad E_n - \Delta_n < E < E_n + \Delta_n;$$

$$g(E) = \frac{m}{\pi\hbar^2}n, \; E_n + \Delta_n < E < E_{n+1} - \Delta_{n+1}. \tag{3.7}$$

Как и в чисто двумерном случае (3.3), здесь плотность состояний дается в расчете на единицу площади. График зависимости (3.7) представлен на рис. 3.1, *д*.

3.2. Статистика носителей в низкоразмерных структурах

Как известно, свойства равновесных электронов в полупроводниках определяются фермиевской функцией распределения и потому зависят от температуры T и положения уровня химического потенциала ζ. Однако в реальных экспериментах известной величиной, как правило, является не ζ, а концентрация носителей. Поэтому одной из первых задач является нахождение связи между уровнем химического потенциала и концентрацией. Мы начнем с нахождения такой связи в двумерном электронном газе, полагая, что все носители находятся в нижней подзоне E_1.

Концентрация носителей n_S находится путем суммирования по всем состояниям подзоны. Согласно (3.2), в интервале энергии dE таких состояний $m dE/(\pi\hbar^2)$, а вероятность заполнения каждого из них дается функцией распределения Ферми $\{\exp[(E-\zeta)/kT]+1\}^{-1}$. Поэтому

$$n_S = \frac{m}{\pi\hbar^2}\int_{E_1}^{\infty}\frac{dE}{\exp[(E-\zeta)/kT]+1} = N_c \ln\{1+\exp[(\zeta - E_1)/kT]\}, \quad (3.8)$$

где двумерная эффективная плотность состояний $N_c = mkT/(\pi\hbar^2)$. Можно преобразовать (3.8), записав ζ как функцию концентрации в явном виде:

$$\zeta = E_1 + kT\ln\left[1+\exp\left(\frac{n_S}{N_c}\right)\right]. \quad (3.9)$$

Используя формулы (3.8) и (3.9), нетрудно получить критерий вырождения двумерного электронного газа и выражения для концентрации в предельных случаях невырожденных и полностью вырожденных электронов. Невырожденный случай реализуется, если

$$n_S \ll \frac{mkT}{\pi\hbar^2}. \quad (3.10)$$

При этом (3.8) принимает более простой вид:

$$n_S = N_c \exp\left(\frac{\zeta - E_1}{kT}\right). \quad (3.11)$$

Формула (3.11) напоминает соответствующее выражение для трехмерного электронного газа, отличаясь лишь сдвигом начала отсчета энергии на E_1 и иным выражением для N_c, более слабо зависящим от температуры ($\sim T$ вместо $\sim T^{3/2}$).

Предел полного вырождения реализуется при выполнении условия, обратного (3.10). В этом пределе

$$\zeta = E_1 + \frac{\pi\hbar^2 n_S}{m}, \quad (3.12)$$

т. е. уровень химического потенциала (энергия Ферми) линейно связан с концентрацией в отличие от трехмерного случая, где $\zeta \sim n^{2/3}$.

Заметим, что в квантово-размерных системах за счет меньшей плотности состояний условие полного вырождения не требует экстремально высоких концентраций или низких температур и достаточно часто реализуется в экспериментах. Например в n-GaAs при $n_S = 10^{12}$ см$^{-2}$ вырождение будет иметь место уже при комнатной температуре.

При выводе формул (3.8)—(3.12) мы предполагали, что носители заполняют лишь одну подзону. Если электронный газ не вырожден, то для этого достаточно выполнения условия (1.5). В случае вырождения необходимо дополнительно, чтобы уровень Ферми лежал ниже E_2. Согласно (3.12), это означает, что

$$n_S < \frac{m}{\pi\hbar^2}(E_2 - E_1). \tag{3.13}$$

Для тонких пленок с энергетическим спектром (1.2) это условие принимает простой вид:

$$n_S a^2 < \frac{3\pi}{2}. \tag{3.14}$$

В квантовых нитях при одной заполненной подзоне линейная концентрация $n = N/L_x$ дается формулой типа (3.8), содержащей под интегралом дополнительный множитель $1/\sqrt{E}$ за счет одномерной плотности состояний (3.5). Такой интеграл в отличие от (3.8) не вычисляется аналитически при произвольном вырождении, и простые формулы могут быть написаны лишь для предельных случаев. В невырожденном одномерном электронном газе формула (3.11) сохраняет свой вид, если под E_1 понимать низший уровень в нити E_{11}, а под N_c — одномерную эффективную плотность состояний $[2mkT/(\pi\hbar^2)]^{1/2}$. В вырожденном случае

$$\zeta = E_{11} + \frac{\pi^2\hbar^2 n^2}{8m}. \tag{3.15}$$

3.3. Примесные состояния и экситоны в низкоразмерных структурах

В главе 3 мы рассмотрели различные методы создания низкоразмерных структур, приводящие к трансформации непрерывного квадратичного спектра свободных носителей в систему дискретных уровней или подзон. Однако потенциальная яма, ограничивая движение носителей узкой областью пространства, изменяет свойства не только свободных, но и связанных электронов. Мы кратко

обсудим влияние размерного квантования на мелкий водородоподобный примесный центр.

Пусть притягивающий центр располагается внутри квантовой ямы в точке с координатой $0 < z_0 < a$. Энергии уровней E_i, создаваемых этим центром, определяются из уравнения Шредингера

$$-\frac{\hbar^2}{2m}\Delta\psi - \frac{e^2}{\kappa\sqrt{x^2+y^2+(z-z_0)^2}}\psi = E_i\psi. \qquad (3.16)$$

Если единственным условием, налагаемым на волновую функцию, является ограниченность, то ответом будет обычная водородоподобная серия уровней $E_i = -\varepsilon_0/i^2$, где $i = 1, 2,...$, а $\varepsilon_0 = me^4/(2\kappa^2\hbar^2)$ — эффективный ридберг. Однако при наличии стенок ямы граничные условия изменяются и для бесконечных стенок принимают вид

$$\psi(0) = \psi(a) = 0. \qquad (3.17)$$

Получить точное решение (3.16) для таких граничных условий в общем виде не удается. Однако это можно сделать в пределе, когда эффективный боровский радиус a_B значительно превосходит ширину потенциальной ямы (толщину пленки):

$$a_B \equiv \frac{\kappa\hbar^2}{me^2} \gg a. \qquad (3.18)$$

При этом в знаменателе кулоновского члена в (3.16) можно опустить член $(z - z_0)^2$, поскольку для электрона на связанном состоянии типичные значения x и y по абсолютной величине имеют порядок a_B, в то время как всюду внутри пленки $|z - z_0| < a$. В результате переменные z и $\rho = (x^2 + y^2)^{1/2}$ в уравнении Шредингера разделяются, что позволяет искать решение (3.16) в виде $\psi = \varphi(\rho)\chi(z)$. Для $\chi(z)$ получается обычное уравнение для частицы в яме с собственными значениями (1.2), в то время как в плоскости xy волновая функция φ должна определяться из уравнения Шредингера для «двумерного атома водорода» с потенциалом $-e^2/\kappa\rho$. Решение этого уравнения можно найти, например, в [30]. Собственные значения энергии для него равны $-\varepsilon_0/(i - 1/2)^2$, т. е. отличаются от случая трехмерной водородоподобной задачи лишь заменой квантового числа i на $i - 1/2$. В результате энергетический спектр примесных состояний в тонкой пленке имеет вид

$$E_i = \frac{\pi^2\hbar^2 n^2}{2ma^2} - \frac{\varepsilon_0}{(i - 1/2)^2}. \qquad (3.19)$$

Условие (3.18) эквивалентно требованию, чтобы получившиеся энергии связи $\varepsilon_0/(i - 1/2)^2$ были значительно меньше энергии размерного квантования. Это означает, что под каждым квантово-раз-

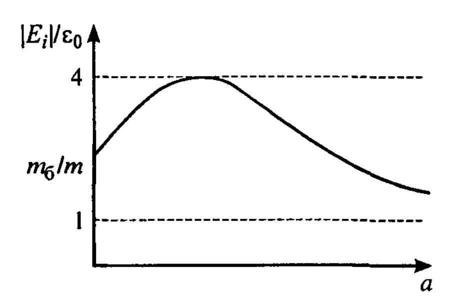


Рис. 3.2. Схематическая зависимость энергии связи примесного центра от ширины квантовой ямы.

мерным уровнем с фиксированным n существует своя ридберговская серия. При этом примесные уровни, связанные со всеми подзонами, кроме основной E_1, находятся на фоне состояний сплошного спектра нижележащих подзон и могут самопроизвольно распадаться за счет переходов в эти состояния. Поэтому указанные уровни имеют конечное время жизни и, следовательно, конечную ширину даже в отсутствие столкновений.

Из (3.19) легко видеть, что основное связанное состояние примесного центра в двумерном случае имеет энергию связи, в 4 раза бóльшую, чем в трехмерном. Физическая картина возрастания энергии связи достаточно понятна. Ограничение стенками ямы приводит к тому, что электрон не может уйти далеко от примеси и в среднем находится от нее на меньшем расстоянии, чем в однородном полупроводнике. Поэтому средняя энергия притяжения, определяющая энергию связи, увеличится.

Граничные условия (3.17) соответствуют модели бесконечно глубокой потенциальной ямы. При конечной глубине ямы зависимость E_i от ширины ямы a может носить более сложный, немонотонный характер. Это связано с тем, что с уменьшением a уровень размерного квантования (при достаточно малых a в яме существует лишь один уровень) становится все более мелким. При этом характерная длина, на которой спадают хвосты волновой функции в широкозонном материале, возрастает и в конечном счете начинает превосходить боровский радиус a_B. Начиная с этого момента размерное квантование перестает оказывать влияние на связанные кулоновские состояния и их энергия связи вновь становится равной ε_0, как это показано на рис. 3.2. Заметим, что в гетероструктурах эффективные массы в материалах ямы (m) и барьера ($m_б$) зачастую различны. При этом предельные значения энергии связи для $a \to \infty$ и $a \to 0$ будут различны. Первая из них определяется величиной ε_0 в узкозонном, а вторая — в широкозонном материале.

Еще более разительно увеличение энергии связи в квантовых нитях [31]. Существенное отличие от только что рассмотренного

случая квантовых ям заключается в невозможности рассмотрения предельного случая нулевой толщины нити. В пленках при $a \to 0$ энергия связи имеет конечное значение $4\varepsilon_0$, а при стремлении к нулю радиуса нити энергия связи стремится к бесконечности. Это означает, что в достаточно тонких нитях примесное состояние может быть в принципе сколь угодно глубоким.

До сих пор при рассмотрении кулоновской задачи как в пленке, так и в нити мы полагали диэлектрическую проницаемость κ одинаковой во всем пространстве. Однако в большинстве реальных структур это далеко не так. Наиболее характерны в этом отношении тонкие пленки, представляющие собой структуры даже не с двумя, а с тремя (подложка—пленка—воздух), порой сильно различающимися проницаемостями. Разница в проницаемостях приводит к тому, что потенциал заряженного центра в структуре, входящий в уравнение Шредингера, уже не носит кулоновский характер, как в (3.16), а существенно искажается за счет сил изображения. Это не может не сказаться на энергии связи примесного центра. Количественные расчеты здесь достаточно сложны (см., например, [32] для тонких пленок и [31] для нитей), но некоторые качественные выводы могут быть легко получены. Если κ в области локализации носителей (в пленке или нити) больше, чем в окружающей среде, то электрическое поле «выталкивается» из указанной области. В результате взаимодействие носителя с полем ослабляется и энергия связи уменьшается по сравнению с диэлектрически однородными системами.

Полученные выше результаты для мелких примесных состояний могут быть непосредственно приложимы для расчета экситонов в квантово-размерных системах. Покажем это для случая двумерной системы. Волновая функция в этом случае зависит от координат как электрона $\mathbf{r}_e$, так и дырки $\mathbf{r}_h$, и уравнение Шредингера более сложно, чем (3.16):

$$-\frac{\hbar^2}{2m_e}\Delta_e\psi - \frac{\hbar^2}{2m_h}\Delta_h\psi -$$

$$-\frac{e^2}{\kappa\sqrt{(x_e - x_h)^2 + (y_e - y_h)^2 + (z_e - z_h)^2}}\psi = E_i\psi. \qquad (3.20)$$

Задача заметно упрощается, если условие (3.18) выполняется для обоих типов носителей. При этом по аналогии с примесным случаем в кулоновском члене (3.20) можно пренебречь слагаемым $(z_e - z_h)^2$. Теперь удобно перейти к новым переменным в плоскости xy:

$$\mathbf{R} = [(m_e x_e + m_h x_h)/(m_e + m_h),\ \ (m_e y_e + m_h y_h)/(m_e + m_h)],$$

$$\rho = (x_e - x_h, y_e - y_h),$$

описывающим соответственно движение центра тяжести системы и относительное движение электрона и дырки. В этих переменных волновая функция системы принимает вид

$$\psi(\mathbf{r}_e, \mathbf{r}_h) = \chi_e(z_e)\chi_h(z_h)\exp(i\mathbf{kR})\varphi(\rho) \tag{3.21}$$

и уравнение Шредингера (3.20) сводится к уравнению для волновой функции относительного движения $\varphi(\rho)$:

$$-\frac{\hbar^2}{2\mu}\Delta_\rho\varphi - \frac{e^2}{\kappa\rho}\varphi = E_{\text{ex}}\varphi, \tag{3.22}$$

где $\mu = m_e m_h/(m_e + m_h)$ — приведенная масса электрона и дырки.

Уравнение (3.22) отличается от уравнения Шредингера для примесных состояний в двумерных системах лишь заменой m на приведенную массу μ. Поэтому ответ для энергии связи экситонов может быть записан по аналогии с (3.19):

$$E_{\text{ex}} = -\frac{m_e m_h e^4}{2\kappa^2\hbar^2(m_e + m_h)(i - 1/2)^2}. \tag{3.23}$$

Энергия связи экситонов, так же как и примесей, в двумерном случае оказывается существенно увеличенной (для основного состояния — в 4 раза). За счет этого экситонные эффекты в квантово-размерных системах оказываются значительно сильнее выражены, чем в массивном образце, и могут наблюдаться при достаточно высоких температурах, включая комнатные.

Рассмотрение экситонных эффектов в квантовых нитях более сложно, чем в квантовых ямах. Показано, однако [31], что с достаточно хорошей точностью вывод о том, что энергия связи экситона отличается от соответствующей энергии для мелкого примесного центра лишь заменой m на μ, сохраняет силу.

В наиболее распространенных полупроводниках, к которым, в частности, принадлежат Si, Ge и соединения III—V, валентная зона носит сложный характер и содержит ветви тяжелых и легких дырок. Поэтому существует два типа экситонов, содержащих соответственно легкую и тяжелую дырку и обладающих в силу этого различными энергиями связи.

Глава 4

ОПТИЧЕСКИЕ СВОЙСТВА КВАНТОВЫХ ЯМ

4.1. Межзонное поглощение

Рассмотрим полупроводник с прямыми разрешенными переходами. В обычном трехмерном случае коэффициент межзонного оптического поглощения такого полупроводника описывается известной формулой [33]:

$$\alpha(\omega)\sim|P_{cv}|^2 \int d^3\mathbf{k}\int d^3\mathbf{k}'\delta\left(E_g+\frac{\hbar^2k^2}{2m_e}+\frac{\hbar^2k^2}{2m_h}-\hbar\omega\right)\int[\Psi^h_{\mathbf{k}}(\mathbf{r})]^*\Psi^e_{\mathbf{k}}(\mathbf{r})d^3\mathbf{r}, \quad (4.1)$$

где P_{cv} — межзонный матричный элемент импульса; E_g — ширина запрещенной зоны; индексы e и h относятся соответственно к электронам и дыркам; $\Psi_{\mathbf{k}}$ — огибающие волновые функции в приближении эффективной массы, представляющие собой плоские волны $\exp(i\mathbf{kr})$; $\delta(x)$ — дельта-функция Дирака.

Мы ограничимся рассмотрением систем типа тонкой пленки или квантовой ямы, в которых эффекты размерного квантования существуют как для электронов, так и для дырок. При этом компонента волнового вектора k_z уже не является хорошим квантовым числом и $\Psi_{\mathbf{k}}(\mathbf{r})$ перестает быть чисто плоской волной:

$$\Psi^e_{\mathbf{k}}(\mathbf{r}) = \exp[i(k_x x + k_y y)]\psi^e_n(z), \quad (4.2)$$

где ψ^e_n — волновая функция квантово-размерного уровня (см., например, (1.1)) и ψ^h_m — аналогичная функция для дырок. В результате интегрирование по k_z и k_z' в (4.1) должно быть заменено на суммирование по номеру квантового уровня. Интегралы по dx и dy дают законы сохранения соответствующих компонент волнового вектора: $\int\exp[i(k_x-k_{x'})]dx = 2\pi\delta(k_x-k_{x'})$ и окончательно

$$\alpha(\omega)=\alpha_0\sum_{m,n}\int[\psi^h_m(z)]^*\psi^e_n(z)\,dz\int\delta(E_g+E^e_n+E^h_m+\varepsilon-\hbar\omega)\,d\varepsilon, \quad (4.3)$$

где α_0 — константа, пропорциональная $|P_{cv}|^2$. Следует подчеркнуть, что, несмотря на резко анизотропный характер двумер-

ных систем, коэффициент межзонного поглощения в них изотропен, т. е. не зависит от поляризации падающего света.

Формула (4.3) содержит интеграл $\int[\psi_m^h(z)]^*\psi_n^e(z)\,dz$, определяющий правила отбора при межзонных переходах. Для бесконечно глубоких квантовых ям волновые функции (1.1) не содержат эффективной массы носителей и потому одинаковы для электронов и дырок. Поэтому в силу ортогональности волновых функций различных состояний данный интеграл отличен от нуля лишь при $n = m$. Для ям конечной глубины правила отбора по-прежнему могут существовать, хотя и не столь жесткие. Если потенциал ямы симметричен (например, для прямоугольной ямы), то функции ψ характеризуются определенной четностью по z, будучи четными при нечетных n и нечетными — при четных. При этом интеграл заведомо обращается в нуль, если m и n имеют различную четность. Непременного равенства $n = m$ здесь не требуется. В асимметричных ямах переходы возможны для любых m и n. Вместе с тем численные расчеты показывают, что при $n \neq m$ интеграл почти всегда значительно меньше единицы. Поэтому в большинстве случаев можно приблизительно считать, что правило отбора для межзонных переходов $n = m$ справедливо для любых ям.

Обсудим вид спектра поглощения в условиях данного правила отбора. Прежде всего отметим, что минимальная энергия поглощаемого фотона равна $E_g + E_1^e + E_1^h$, т. е. энергия края поглощения больше, чем в однородном полупроводнике, и растет с уменьшением ширины квантовой ямы. В интервале энергий фотона $E_g + E_1^e + E_1^h < \hbar\omega < E_g + E_2^e + E_2^h$ коэффициент поглощения постоянен, а при $\hbar\omega = E_g + E_2^e + E_2^h$ скачком увеличивается. В целом $\alpha(\omega)$ имеет ступенчатый вид со скачками при $\hbar\omega = E_g + E_n^e + E_n^h$. Это есть отражение известного утверждения теории полупроводников о том, что при прямых разрешенных оптических переходах спектр поглощения воспроизводит функцию плотности состояний. Если правило отбора $n = m$ не выполняется, то возникает дополнительное дробление $\alpha(\omega)$ на более мелкие ступеньки, отвечающие $\hbar\omega = E_g + E_n^e + E_m^h$ с $n \neq m$.

Сделанный вывод о виде спектра поглощения получен в идеализированной модели невзаимодействующих носителей. Экситонные эффекты могут заметно исказить $\alpha(\omega)$. Во-первых, вблизи каждой ступеньки появляется серия экситонных пиков. Во-вторых, кулоновский потенциал электронно-дырочного взаимодействия не только создает дискретные связанные состояния, но и искажает плотность состояний в непрерывном спектре на расстоянии $\sim\varepsilon_0$ от края. Поэтому $\alpha(\omega)$ отличается от прямоугольных ступенек как слева, так и справа от них.

Строгое теоретическое рассмотрение [30] дает следующую формулу для спектра поглощения вблизи края с учетом экситонных эффектов:

$$\alpha(\omega)=\alpha_0\left[\sum_{i=1}^{\infty}\frac{4}{(i-1/2)^3}\delta\left(\Delta+\frac{1}{(i-1/2)^2}\right)+\Theta(\Delta)\frac{\exp(\pi/\sqrt{\Delta})}{\mathrm{ch}(\pi/\sqrt{\Delta})}\right], \quad (4.4)$$

где $\Delta = (\hbar\omega - E_g - E_1^e - E_1^h)/\varepsilon_0$, а $\Theta(x)$ — единичная функция, описанная в разделе 3.1. Первый член в (4.4) описывает систему экситонных пиков, а второй — искажение края непрерывного поглощения. Фактор $\exp(\pi/\sqrt{\Delta})/\mathrm{ch}(\pi/\sqrt{\Delta})$ стремится к двум при $\Delta \to 0$ и к единице при $\Delta \to \infty$. Это означает, что на краю поглощения α двукратно возрастает по сравнению с идеальной ступенькой (рис. 4.1).

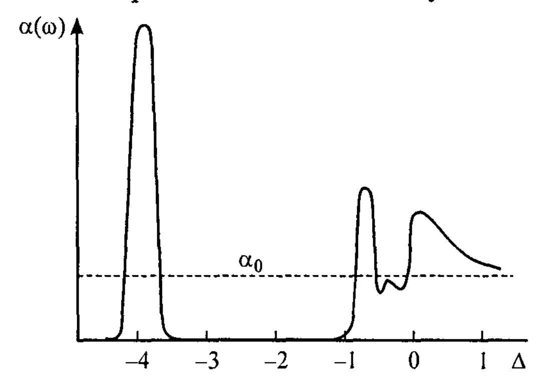


Рис. 4.1. Спектр межзонного поглощения двумерной системы с учетом экситонных эффектов ($\Delta = (\hbar\omega - E_g - E_1^e - E_1^h)/\varepsilon_0$).

В разделе 3.4 уже упоминалось о наличии двух типов экситонов (с легкой и тяжелой дыркой) во всех важнейших полупроводниках, используемых для изготовления квантовых структур. Это приведет к дополнительному двукратному расщеплению экситонных пиков, показанных на рис. 4.1, и аналогичных пиков вблизи краев вышележащих зон.

4.2. Межуровневые переходы

Свет с частотами, меньшими края межзонного поглощения, рассмотренного в предыдущем параграфе, может вызывать лишь переходы внутри энергетических зон. Такое внутризонное поглощение в отличие от межзонного существенно различно для света разной поляризации. Если свет падает по нормали к двумерному слою, то электрическое поле световой волны лежит в плоскости слоя, где электроны ведут себя как свободные. Поэтому свет с такой поляризацией испытывает лишь обычное поглощение на свободных но-

сителях, существующее лишь в меру их рассеяния, имеющее небольшую величину и монотонную зависимость от частоты света.

Значительно более интересно поведение света, поляризованного перпендикулярно плоскости слоя. Такой свет может вызывать резонансное поглощение между уровнями размерного квантования. Наличие или отсутствие поглощения (правила отбора) и его относительная интенсивность определяются межуровневым матричным элементом импульса

$$P_{mn} = -i\hbar \int \psi_m^*(z) \frac{d\psi_n(z)}{dz} dz. \qquad (4.5)$$

Для симметричных квантовых ям правила отбора противоположны тем, что имеют место для межзонных переходов (см. предыдущий раздел): межуровневые переходы возможны лишь между состояниями противоположной четности, иными словами, электроны с первого уровня могут переходить лишь на второй, четвертый и др., но не на третий или пятый. Если же яма не обладает свойством четности, как например в МДП-структуре или одиночном гетеропереходе, то оптические переходы могут происходить между любыми уровнями.

Из сказанного следует, что спектр внутризонного поглощения z-поляризованного света представляет собой ряд узких резонансных полос. Если в квантовой яме имеется N уровней и заполнен только нижний из них, то таких полос будет $N-1$ в несимметричной и $[N/2]$ в симметричной яме (выражение [...] означает целую часть числа). В МДП-структурах наблюдают такое поглощение и убеждаются в его резонансном характере чаще всего не путем изменения частоты света ω, а путем изменения напряжения на затворе при фиксированной ω [34]. При этом меняются энергетические зазоры между уровнями, определяемые величиной электрического поля затвора, и в какой-то момент для данной ω реализуется условие резонанса.

Сделанный выше вывод о том, что описанное резонансное поглощение возможно только для света, поляризованного по нормали к квантовым слоям, справедлив для изотропного закона дисперсии носителей (например, в n-GaAs) или же для определенной симметричной ориентации слоев (например, в МДП-структурах на поверхности (100) кремния). При наличии анизотропии межуровневые переходы могут вызываться и светом с поляризацией, параллельной слоям. Это относится, в частности, к дыркам в валентной зоне наиболее популярных полупроводников Si, Ge, GaAs, InP и др. [34].

4.3. Оптическая ионизация квантовых ям

Обратимся к рассмотрению внутризонных оптических свойств квантовых ям конечной глубины, где наряду с дискретными уровнями E_n существует континуум делокализованных состояний с энерги-

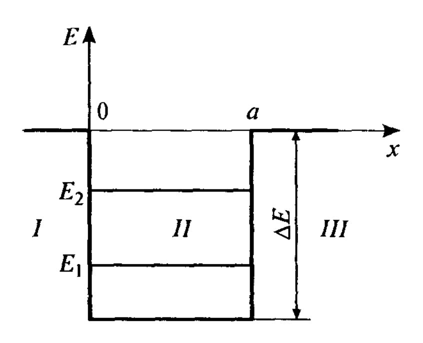


Рис. 4.2. Энергетический профиль квантовой ямы.

ями выше края ямы. При этом свет, поляризованный перпендикулярно слоям и вызывающий при $\hbar\omega = E_n - E_m$ переходы между квантово-размерными уровнями, может при больших частотах, $\hbar\omega > \Delta E - E_1$ (ΔE — глубина ямы), вызывать переходы с основного уровня в континуум, т. е. производить фотоионизацию квантовых ям.

Для вычисления матричного элемента, определяющего вероятность таких переходов, нам потребуется знание волновых функций континуума ψ_ε. Для прямоугольной ямы, изображенной на рис. 4.2, расчет ψ_ε весьма прост и имеется в любом задачнике по квантовой механике. Волновая функция электрона с энергией ε может быть записана в виде

$$\begin{aligned} \psi_\varepsilon &= \exp(ikx) + A\exp(-ikx) && \text{в области } I, \\ \psi_\varepsilon &= B\exp(iKx) + C\exp(-iKx) && \text{в области } II, \\ \psi_\varepsilon &= D\exp(ikx) && \text{в области } III, \end{aligned} \tag{4.6}$$

где $k^2 = 2m\varepsilon/\hbar^2$, $K^2 = 2m(\Delta E + \varepsilon)/\hbar^2$. Коэффициенты A, B, C, D определяются из условий сшивки на границах ямы. Например, для коэффициента B, определяющего амплитуду волновой функции непосредственно над ямой, имеем [35]:

$$B = \frac{2(K/k+1)}{(K/k+1)^2 - (K/k-1)^2\exp(2iKa)}. \tag{4.7}$$

Видно, что для медленных электронов с $\varepsilon \ll \Delta E$ $B \sim k$, т. е. очень мало́. Аналогичной малостью обладают коэффициенты C и D. Такие электроны мало проникают в непосредственную окрестность квантовой ямы, а почти полностью отражаются от нее. Исключение составляют так называемые резонансные ямы, для которых

$$\frac{\sqrt{2m\Delta E}}{\pi\hbar} = 1,2,3,\ldots \tag{4.8}$$

Для этих ям $B \approx C \approx 1/2$, $D \approx 1$, т. е. электроны свободно проходят над ямой, не отражаясь. Соотношение (4.8) есть условие того, что

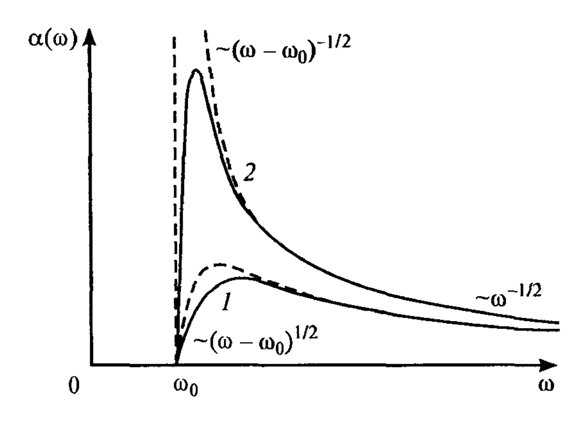


Рис. 4.3. Схематический вид спектров фотоионизации нерезонансной (*1*) и резонансной (*2*) квантовых ям. *Штриховые кривые* — простая модель; *сплошные* — с учетом деполяризационных эффектов (см. раздел (4.4)). $\omega_0 = (\Delta E - E_1)/\hbar$ — пороговая частота фотоионизации.

квантовая яма представляет собой как бы резонатор Фабри—Перо для электронных волн де Бройля.

Из сказанного следует, что спектры оптической ионизации резонансных и нерезонансных квантовых ям должны существенно различаться. Действительно, при частотах света, близких к порогу ионизации $\hbar\omega_0 = \Delta E - E_1$, электроны выбрасываются в состояния с малыми энергиями $\varepsilon = \hbar\omega - \Delta E + E_1$. В окрестности нерезонансной ямы (в области *II*) волновая функция таких электронов имеет малую амплитуду, в то время как волновая функция первого уровня $\psi_1(z)$ локализована именно в этой области. Поэтому любые матричные элементы между ψ_ε и ψ_1 будут малы и коэффициент поглощения вблизи порога обращается в нуль. Строгие расчеты [36] показывают, что $\alpha(\omega) \sim (\hbar\omega - \Delta E + E_1)^{1/2}$. Для резонансных ям, напротив, матричный элемент в припороговой области не мал и коэффициент поглощения не только не зануляется, но неограниченно возрастает вблизи порога: $\alpha(\omega) \sim (\hbar\omega - \Delta E + E_1)^{-1/2}$. При больших частотах света ω электроны выбрасываются в высоколежащие состояния, практически не возмущенные потенциалом ямы. Свойства электронов в этих состояниях близки к свойствам свободных электронов и коэффициент поглощения стремится к нулю. Характерный вид спектров оптической ионизации квантовых ям показан на рис. 4.3.

4.4. Эффекты деполяризации

Выводы предыдущих разделов 4.2 и 4.3 о характере спектров внутризонного поглощения в квантово-размерных системах получены без учета так называемых эффектов деполяризации. Суть

этих эффектов заключается в том, что для света, поляризованного по нормали к слоям, высокочастотное электрическое поле должно удовлетворять условию непрерывности электрической индукции. Поэтому поле, реально действующее на электроны, отличается от поля падающей световой волны на множитель, равный диэлектрической проницаемости квантово-размерной системы κ. Вблизи частот, отвечающих сильному (например, межуровневому) поглощению, κ резко зависит от частоты света. Это обстоятельство может привести с заметному искажению спектров поглощения $\alpha(\omega)$.

Для количественного описания эффектов деполяризации необходимо прежде всего иметь правильное выражение для функции $\kappa_{zz}(\omega)$, описывающей частотную дисперсию диэлектрической проницаемости системы для нормально поляризованного света. Это выражение будет различно для частот, отвечающих межуровневым переходам (см. раздел 4.2), и частот, превосходящих порог оптической ионизации квантовых ям (см. раздел 4.3). Обсудим вначале первый случай, ограничиваясь для наглядности простой моделью периодической системы, содержащей ямы шириной a с концентрацией двумерных носителей n_S и барьеры шириной b, не содержащие носителей [1].

Для света с z-поляризацией и частотой, близкой к частоте межуровневого резонанса $\omega_{12} = (E_2 - E_1)/\hbar$, квантовая яма может рассматриваться как двухуровневая электронная система. Диэлектрическая проницаемость такой системы, содержащая как решеточную, так и электронную компоненту, может быть записана в следующем виде [4]:

$$\kappa_1 = \kappa_0 + \frac{4\pi n_S e^2}{ma} \frac{1}{\omega_{12}^2 - \omega^2 - 2i\omega/\tau}, \tag{4.9}$$

где κ_0 — диэлектрическая проницаемость решетки и заполненных электронных оболочек, а τ — время релаксации носителей. Область барьера не содержит свободных носителей и имеет проницаемость $\kappa_2 = \kappa_0$.

Получим выражение для эффективной проницаемости такой слоистой системы. Пусть F_1 и F_2 — напряженности электрического поля световой волны в материалах ямы и барьера. Для z-поляризованного света они связаны условием непрерывности электрической индукции: $\kappa_1 F_1 = \kappa_2 F_2$. Поэтому эффективная диэлектрическая проницаемость, связывающая средние значения индукции и поля, равна:

$$\kappa_{zz} = \frac{aD_1 + bD_2}{aF_1 + bF_2} = \frac{D_1(a+b)}{aD_1/\kappa_1 + bD_1/\kappa_2} = \frac{\kappa_1\kappa_2(a+b)}{a\kappa_2 + b\kappa_1}. \tag{4.10}$$

Подставляя в (4.10) указанные выше значения κ_1 и κ_2, получаем:

$$\kappa_{zz}(\omega) = \kappa_0 \frac{\omega_{12}^2 - \omega^2 - 2i\omega/\tau}{\omega_{12}^2 + \omega^2 - \omega_p^2 - 2i\omega/\tau}, \qquad (4.11)$$

где $\omega_p^2 = 4\pi n_S e^2 b/[m(a+b)]$.

Видно, что мнимая часть κ_{zz}, описывающая коэффициент поглощения, имеет резонансный пик при $\omega = (\omega_{12}^2 + \omega_p^2)^{1/2}$. Это означает, что за счет эффектов деполяризации частота резонансного межуровневого поглощения сдвигается на величину, определяемую эффективной плазменной частотой ω_p.

Сходным образом может быть рассмотрено влияние деполяризационных эффектов и на спектр оптической ионизации квантовых ям [37, 38]. В отличие от межуровневых переходов, где эти эффекты сдвигают линию поглощения, здесь положение порога ионизации остается неизменным, хотя форма спектра и изменяется. Основной результат, как и для межуровневых переходов, сводится к резкому подавлению поглощения там, где в отсутствие деполяризации оно было бы очень велико. При этом для нерезонансных ям спектр поглощения существенно не изменится, а для резонансных расходимость коэффициента поглощения на пороге ионизации сменится обращением в нуль. В результате квантовые ямы с параметрами, близкими к резонансным, будут иметь спектр поглощения с резким максимумом вблизи порога ионизации (сплошная кривая *1* на рис. 4.3).

Глава 5

КИНЕТИЧЕСКИЕ ЭФФЕКТЫ В ДВУМЕРНЫХ СИСТЕМАХ

Обсудим некоторые черты проводимости и других кинетических явлений в слое двумерного электронного газа. Поскольку в плоскости слоя носители ведут себя как свободные, их поведение в этой плоскости можно описать в рамках классического кинетического уравнения аналогично тому, как это делается для обычного трехмерного электронного газа.[1] Отличие от трехмерного случая может возникнуть, во-первых, за счет иной плотности состояний в двумерной системе (см. раздел 3.1) и, во-вторых, благодаря наличию в двумерных структурах некоторых специфических механизмов рассеяния, отсутствующих в трехмерном электронном газе. Рассмотрим последовательно оба этих вопроса.

5.1. Время релаксации и подвижность

Чтобы понять, как соотносятся между собой подвижности носителей в двумерном и трехмерном электронном газе, сопоставим между собой процессы рассеяния для обоих случаев. В свободном трехмерном газе волновые функции электронов представляют собой плоские волны. В двумерных системах рассеяние связано с изменением импульса в плоскости слоя, где волновые функции — также плоские волны. Поэтому для одного и того же потенциала рассеяния $V(\mathbf{r})$ матричный элемент рассеяния $M(\mathbf{q})$ ($\mathbf{q}$ — переданный импульс) в дву- и трехмерном случаях будет одинаков.

Время релаксации τ, определяющее подвижность носителей, связано с $M(\mathbf{q})$ известной формулой [39]:

$$\frac{1}{\tau}=\frac{2\pi}{\hbar}\sum_q |M(\mathbf{q})|^2\delta\left[\frac{p^2}{2m}-\frac{(\mathbf{p}-\mathbf{q})^2}{2m}\right]\frac{\mathbf{p}\cdot\mathbf{q}}{p^2}. \tag{5.1}$$

Различие между дву- и трехмерным случаями состоит в том, что суммирование (т. е. интегрирование) в (5.1) ведется в первом слу-

[1] Исключение представляет случай сильных магнитных полей, требующий специфически квантового рассмотрения, которое будет проведено в следующей главе.

чае по двумерному, а во втором по трехмерному пространству возможных переданных импульсов **q**. В первом случае $\sum_q \to \frac{S}{4\pi^2\hbar^2}\int q dq \int q\theta$, а во втором $\sum_q \to \frac{V}{4\pi^2\hbar^2}\int q^2 dq \int \sin\theta d\theta$, где θ — угол между векторами **q** и **p**, а V и S — соответственно объем и площадь образца.

Для важнейших механизмов рассеяния $M(\mathbf{q})$, как правило, является степенной функцией переданного импульса: $|M|^2 = Aq^S$ [1]. При этом вычисления по формуле (5.1) дают для двумерного случая $\tau_2^{-1} \sim mSAp^S/\hbar^3$, а для трехмерного $\tau_3^{-1} \sim mVAp^{S+1}/\hbar^4$ (мы не писали точные численные множители, возникающие при вычислении соответствующих интегралов). Поскольку объем пленки равен произведению ее площади на толщину a, то $V/S = a$, и из сравнения этих формул получаем:

$$\tau_2/\tau_3 = \gamma\, pa/\hbar, \tag{5.2}$$

где γ — численный множитель порядка единицы.

Входящий в формулу (5.2) импульс носителя p имеет порядок теплового импульса $\sqrt{m\kappa T}$ для невырожденного случая и импульса Ферми $\hbar\sqrt{2\pi n}$ для вырожденного. Поэтому условия заполнения одной подзоны (1.6) и (3.13) эквивалентны требованию, чтобы правая часть (5.2) была меньше единицы. Таким образом, при одном и том же механизме рассеяния время релаксации в двумерном газе меньше, чем в массивном полупроводнике, и уменьшается с уменьшением толщины двумерного слоя.

Другим отличием τ_2 от τ_3 является его иная зависимость от энергии электрона (т. е. импульса p). Благодаря этому в невырожденном случае, когда $p \sim \sqrt{T}$, время релаксации для двумерного газа приобретает дополнительный множитель $T^{1/2}$, изменяющий температурную зависимость подвижности μ. В частности, при рассеянии на акустических фононах вместо зависимости $\mu \sim T^{-3/2}$, характерной для массивных полупроводников, в двумерном случае имеем $\mu \sim T^{-1}$, а при рассеянии на ионах примеси вместо $\mu \sim T^{3/2}$ имеем $\mu \sim T^2$. В вырожденном случае соотношение (5.2) приводит к иной, нежели в трехмерном случае, зависимости μ от импульса Ферми, т. е. от концентрации носителей.

5.2. Механизмы рассеяния

В предыдущем разделе мы обсудили некоторые общие свойства подвижности двумерных систем и сопоставили значения двумерной и трехмерной подвижностей при одном и том же рассеиваю-

щем потенциале. Однако в двумерных системах сами механизмы рассеяния могут отличаться от механизмов для массивных образцов. Мы рассмотрим четыре механизма рассеяния, играющих основную роль в двумерных системах: рассеяние на ионизованных примесях, фононное рассеяние, сплавное рассеяние и рассеяние на стенках квантовой ямы. Последнее из них специфично только для двумерных систем, а три других характерны и для массивных образцов, но в двумерном случае имеют свою специфику и потому также нуждаются в отдельном рассмотрении.

Рассеяние на ионизованных примесях. Основное отличие от трехмерного случая здесь состоит в том, что рассеивающие центры зачастую располагаются не в плоскости двумерного газа, а на некотором расстоянии от нее. В МДП-структурах — это ионы так называемого встроенного заряда, неизменно существующего в оксиде [2], а в гетероструктурах с модулированным легированием (см. раздел 5.3) — ионы легирующей примеси, намеренно удаленные от электронного газа. Поэтому подвижность здесь будет определяться не только концентрацией примеси, но и пространственным расположением ионов.

Рассмотрим матричный элемент рассеяния с изменением импульса q в первой подзоне двумерной системы, вызванного кулоновским центром, расположенным в точке $(0, 0, z_i)$:

$$M(q) = \frac{e^2}{\kappa S}\int\int\int\frac{\exp[i(q_x x/\hbar + q_y y/\hbar)]}{[x^2+y^2+(z-z_i)^2]^{1/2}}\psi_1^2(z)\,dx\,dy\,dz =$$

$$= \frac{2\pi e^2}{\kappa S}\int\frac{I_0(q\rho/\hbar)\rho d\rho}{[\rho^2+(z-z_i)^2]^{1/2}}\psi_1^2(z)\,dz =$$

$$= \frac{2\pi\hbar e^2}{\kappa q S}\int\exp(-|z-z_i|q/\hbar)\psi_1^2(z)\,dz, \qquad (5.3)$$

где $I_0(x)$ — функция Бесселя нулевого порядка. В ходе вычислений мы использовали следующие формулы [40]:

$$\int_0^{2\pi}\exp(ia\cos\varphi)\,d\varphi = 2\pi I_0(a), \qquad \int_0^{\infty}\frac{I_0(bx)}{\sqrt{x^2+c^2}}x\,dx = b^{-1}\exp(-|c|b).$$

Выражение (5.3) может быть упрощено. Очевидно, что изменение импульса при упругом рассеянии q не может превосходить удвоенный импульс электрона $2p$.[2] Как уже отмечалось выше, в условиях заполнения одной подзоны $pa/\hbar < 1$. Поэтому в интервале значений $0 < z < a$, в котором $\psi_1(z)$ имеет заметную величину,

[2] С формальной точки зрения это следует из закона сохранения энергии, стоящего в (5.1) в аргументе δ-функции.

$\exp(-|z-z_i|q/\hbar)$ меняется слабо и может быть вынесена из-под интеграла в (5.3). В результате получаем:

$$M(q)=\frac{2\pi\hbar e^2}{\kappa qS}\exp(-|z_i|q/\hbar). \tag{5.4}$$

Подставляя (5.4) в (5.1) и интегрируя по углу θ, получаем

$$\frac{1}{\tau}=\frac{4\pi me^4}{\kappa^2\hbar p^2}N_i\int_0^1\exp(-4|z_i|px/\hbar)\frac{dx}{\sqrt{1-x^2}}, \tag{5.5}$$

где N_i — поверхностная концентрация рассеивателей. Появление этого множителя имеет следующую причину. Матричный элемент $M(q)$ описывает рассеяние на одиночной примеси. При вычислении τ^{-1} вероятность такого рассеяния необходимо умножить на полное число рассеивателей, которое в комбинации с фактором S^{-1}, содержащимся в (5.4), дает N_i.

Стоящий в (5.5) интеграл при малых z_i равен $\pi/2$, а при больших z_i стремится к $(4|z_i|p)^{-1}$. В целом формула (5.5) отражает основные закономерности примесного рассеяния в двумерных системах: время релаксации (т. е. подвижность) растет с ростом кинетической энергии двумерных электронов $p^2/2m$ и с увеличением расстояния между двумерным газом и рассеивающими центрами.

Рассеяние на фононах. При рассмотрении рассеяния двумерных электронов на фононах чаще всего полагают, что фононный спектр системы остается таким же, как и в массивном полупроводнике. Это позволяет применять выражения, полученные в разделе 5.1. При рассеянии на акустических фононах $s=0$ и, кроме того, μ содержит дополнительный множитель T^{-1}, отражающий температурную зависимость числа фононов в образце [39]. В результате для двумерного газа независимо от его вырождения $\mu\sim T^{-1}$ и не зависит от n_S.[3]

Сплавное рассеяние. Этот своеобразный механизм имеет место только в полупроводниковых твердых растворах. Если взять микроскопический локальный объем в таком веществе, то за счет статистического характера распределения атомов веществ, образующих твердый раствор, отношение концентраций его компонент, т. е. локальный состав раствора в этом микрообъеме, может отличаться от среднего по образцу. Поскольку ширина запрещенной зоны E_g определяется составом твердого раствора, то в образце возникают флуктуации E_g и, следовательно, случайный потенциал, модулирующий

[3] Последнее утверждение несправедливо для МДП-структур. В них концентрация n_S определяет не только импульс Ферми, но и эффективную ширину ямы a, входящую в (5.2). За счет этого появляется слабая зависимость $\mu\sim n_S^{-1/3}$.

энергетические зоны. Этот потенциал и вызывает дополнительное рассеяние носителей.

В отличие от кулоновского потенциала заряженных примесей описанный выше сплавной потенциал является короткодействующим. Он существует только в непосредственной близости от узла решетки, где вместо атома одного вещества стоит атом другого. Формально это означает, что рассеивающий потенциал $V(\mathbf{r})$ может быть записан в виде суммы δ-функций. Интенсивность рассеяния с изменением импульса на величину $\mathbf{q}$ пропорциональна соответствующей фурье-компоненте $V(\mathbf{r})$. Точечный δ-потенциал характерен тем, что его компоненты $V_{\mathbf{q}} = \text{const}(\mathbf{q})$ [41]. Поэтому сплавное рассеяние с любым $\mathbf{q}$ равновероятно, т. е. $|M| = \text{const}(\mathbf{q})$ или $s = 0$. Аналогичное условие имело место для рассеяния на акустических фононах (см. выше), и потому, как и там, подвижность не зависит от n_S в пленках и квантовых ямах и падает пропорционально $n_S^{-1/3}$ в инверсионных каналах МДП- и гетероструктур [42].

Сплавное рассеяние наиболее заметно в гетероструктурах, где узкозонный материал, в котором сосредоточены носители, является твердым раствором. Однако некоторые эффекты могут сохраняться и в случае, когда узкозонный полупроводник — чистое вещество, окруженное широкозонным твердым раствором (как в системе GaAs–AlGaAs). Если глубина квантовой ямы не слишком велика, то волновые функции ψ_N имеют «хвосты» в широкозонном материале, которые и обусловливают существование сплавного рассеяния в этом случае [43].

Рассеяние на стенках квантовой ямы. Данный механизм рассеяния обусловлен неидеальностью потенциальных стенок, ограничивающих движение носителей в квантовой яме. Роль таких стенок в тонких пленках играют внешние поверхности пленки, в МДП-структурах — граница полупроводник—диэлектрик, а в гетероструктурах — граница узкозонного и широкозонного полупроводников. Если указанные границы не являются идеальными плоскостями, а содержат шероховатости, то отражение носителей от них не является абсолютно зеркальным и приводит к частичной потере направленного импульса носителей, т. е. вызывает релаксацию импульса.

Матричный элемент рассматриваемого рассеяния может быть получен без труда. Пусть потенциальный барьер на рассеивающей границе имеет высоту V_0, а сама граница, которая в идеальном случае представляла бы плоскость $z = 0$, при наличии шероховатостей описывается формулой $z = \Delta(x, y)$. Используя понятие единичной функции $\Theta(x)$, введенное в разделе 3.1, можем записать потенциал границы в виде $V_0\Theta[-z + \Delta(x, y)]$. Его отличие от потенциала идеальной границы $\Delta V = V_0\,\Theta[-z + \Delta(x, y)] - V_0\,\Theta(-z)$ при малых Δ

приблизительно равно $V_0\,\Delta(x, y)\,\delta(z)$, где $\delta(z)$ — δ-функция Дирака. В результате искомый матричный элемент имеет вид

$$M(q) = \frac{1}{S}\int\int\int \exp[i(q_x x + q_y y)]V_0\,\Delta(x,y)\delta(z)\left|\psi_1(z)\right|^2 dx dy dz =$$

$$= \Delta_q V_0 \left|\psi_1(0)\right|^2, \qquad (5.6)$$

где Δ_q — фурье-компонента $\Delta(x, y)$.

Если барьер на гетерогранице очень велик ($V_0 \to \infty$), то волновая функция на границе стремится к нулю и (5.6) содержит неопределенность типа $0 \cdot \infty$. Чтобы ее раскрыть, учтем, что в подбарьерной области $z > 0$ волновая функция ψ_1 имеет экспоненциальный вид: $\psi_1(z) \cong \cong \psi_1(0)\exp[-(2mV_0/\hbar^2)^{1/2}z]$. Поэтому $\psi_1'(0) = -\psi_1(0)(2mV_0/\hbar^2)^{1/2}$, что позволяет переписать (5.6) в виде

$$M(q) = \frac{\hbar^2}{2m}\left|\psi_1'(0)\right|^2 \Delta_q\,. \qquad (5.7)$$

Видно, что интенсивность рассеяния на границе определяется как амплитудой шероховатостей с данным периодом, так и производной волновой функции на границе. Последний фактор определяется шириной квантовой ямы a. Поскольку волновая функция сосредоточена в области с размерами $\sim a$ и нормирована на единицу, то амплитуда ψ_1 имеет порядок $a^{-1/2}$. При этом $\left|\psi_1'\right| \sim a^{-3}$, т. е. интенсивность рассеяния на границе резко возрастает с уменьшением ширины квантовой ямы, а в случае МДП-структуры — с увеличением напряжения на затворе.

Рассмотренный механизм рассеяния обычно не существен в гетероструктурах, где при выполнении условий, сформулированных в разделе 2.5, качество границ, как правило, весьма высоко. Вмес-

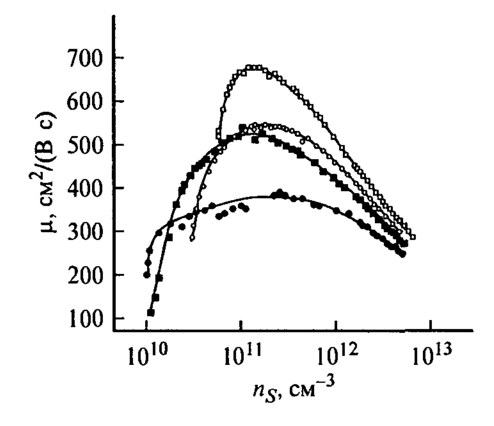


Рис. 5.1. Зависимость подвижности носителей от их концентрации n_S (т. е. от напряжения на затворе V_3 (см. (1.9)) для нескольких различных МДП-структур [4]).

те с тем рассеяние на границе в МДП-структурах достаточно интенсивно и играет наряду с рассеянием на заряженных центрах в оксиде определяющую роль в подвижности носителей μ в инверсионных каналах. Совместное действие этих двух механизмов приводит к своеобразной зависимости μ от напряжения на затворе V_g (т. е. от концентрации носителей в канале n_S). При рассеянии на заряженных центрах подвижность растет с ростом V_g, так как при этом растет энергия Ферми, а следовательно и импульс p, что, согласно (5.5), уменьшает рассеяние. При рассеянии на шероховатостях, как только что отмечалось, подвижность с ростом V_g должна падать. В результате зависимость $\mu(V_g)$ представляет собой кривую с максимумом (рис. 5.1), где область начального роста μ отвечает рассеянию на зарядах в оксиде, а область последующего спада — рассеянию на границе полупроводник—диэлектрик.

5.3. Модулированное легирование

Проведенный в предыдущем разделе анализ механизмов рассеяния позволяет понять, что необходимо сделать для получения максимально высокой подвижности в двумерном электронном газе. Прежде всего следует иметь высокое качество стенок потенциальной ямы, чтобы исключить рассмотренное в разделе 5.2 поверхностное рассеяние. Как уже говорилось в разделе 2.3, из всех рассмотренных типов квантово-размерных систем это условие лучше всего достигается в гетероструктурах. При этом для подавления сплавного рассеяния предпочтительны структуры, в которых узкозонная компонента представляет собой чистое соединение, а не твердый раствор (к таким структурам, в частности, относится и GaAs–$Al_xGa_{1-x}As$). Путем понижения температуры подавляется фононное рассеяние, и остается решить задачу сведения к минимуму рассеяния на заряженных примесях.

Сразу отметим, что для вырожденного электронного газа тривиальный, на первый взгляд, вывод о необходимости снижения для этой цели концентрации легирующих примесей N_t оказывается неверен. Действительно, понижение N_t приводит, в силу нейтральности образца, к понижению концентрации электронов. Это в свою очередь уменьшает импульс Ферми и, согласно (5.5), приводит к понижению подвижности, в точности компенсирующему ее рост за счет уменьшения концентрации рассеивающих центров.[4] Остается

[4] Подчеркнем, что речь идет о концентрации намеренно вводимых легирующих примесей. Существующие в образце фоновые примеси частично носят компенсирующий характер. Концентрацию таких примесей следует всяческн понижать, так как при этом снижается число рассеивателей и одновременно растут концентрации носителей, импульс Ферми, а следовательно, и подвижность.

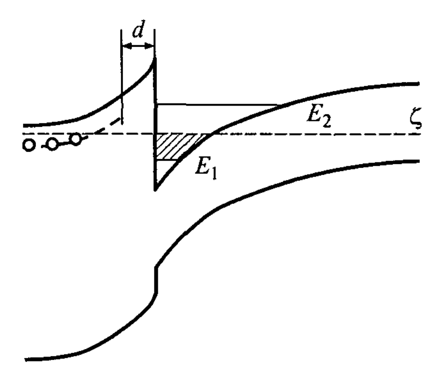


Рис. 5.2. Зонная диаграмма гетероструктуры с модулированным легированием. В широкозонном материале показаны заполненные и ионизованные примесные состояния. Заштрихована область, содержащая двумерный электронный газ.

один путь для роста подвижности — увеличивать расстояние z_i между носителями и примесями.

Для решения этой задачи была предложена концепция модулированного легирования гетероструктур [13]. Схема одиночной гетероструктуры с модулированным легированием показана на рис. 5.2. Основная идея заключается в том, что в гетеропереходе легируется широкозонный полупроводник, а узкозонный оставляется нелегированным. Чтобы уравнять уровни химического потенциала в обоих полупроводниках, некоторое количество носителей переходит в узкозонный полупроводник, образуя вблизи границы слой электронного (дырочного — при *p*-легировании) газа. В результате ионизованные примеси и свободные носители оказываются по разные стороны гетерограницы. Подобное пространственное разделение и приводит к увеличению подвижности.

Еще большего возрастания подвижности можно достичь, вводя в структуру так называемый спейсер, т. е. оставляя нелегированным тонкий слой широкозонного материала толщиной *d*, непосредственно примыкающий к гетерогранице (рис. 5.2). Этим достигаются еще большее разделение носителей и рассеивающих центров и дальнейшее увеличение подвижности. Путем увеличения *d* до значений порядка 800 Å в структурах GaAs–AlGaAs было получено рекордное значение подвижности электронов, при низкой температуре превосходящее 10^7 см^2/(В · с).

Однако увеличивать подвижность носителей путем увеличения толщины спейсера можно лишь до определенного предела. Дело в том, что рост *d* приводит к уменьшению концентрации двумерных носителей, что в свою очередь уменьшает их энергию Ферми. А поскольку, согласно (5.5), уменьшение энергии носителей понижает их подвижность, данный эффект приводит к наличию максиму-

ма на зависимостях $\mu(d)$ и делает нецелесообразным дальнейшее увеличение толщины спейсера.

С помощью модулированного легирования можно повышать подвижность двумерных носителей не только в одиночных гетероструктурах, но и в квантовых ямах. Однако при этом получаются значения подвижности, существенно меньшие указанных выше рекордных величин. Причина кроется в особенностях эпитаксиального роста в системе GaAs–AlGaAs. Многочисленные опыты показывают, что качество гетероструктуры GaAs–AlGaAs, полученной путем роста твердого раствора на арсениде галлия (прямая гетероструктура), значительно выше, чем при обратной последовательности роста (инверсная гетероструктура). Рекордные значения подвижности могут быть получены лишь на прямых гетероструктурах. Поскольку в квантовой яме одна из гетерограниц неизбежно инверсионная, рассеяние на ней снижает подвижность по сравнению со случаем одиночной прямой гетероструктуры.

5.4. Межуровневое рассеяние

До сих пор при описании кинетических явлений мы считали, что носители заполняют лишь первую подзону размерного квантования, т. е. могут рассматриваться как чисто двумерный газ. Вместе с тем в условиях, когда требование (3.13) нарушается и энергия носителей превосходит энергию второго уровня E_2, появляется ряд интересных эффектов, связанных с межуровневым (межподзонным) рассеянием.

Рассмотрим носитель, находящийся в первой квантово-размерной подзоне и имеющий в ней кинетическую энергию $(p_x^2 + p_y^2)/2m > (E_2 - E_1)$. Тогда помимо процессов упругого рассеяния, описанных в разделах 5.1 и 5.2 и связанных с поворотом вектора импульса в плоскости xy, он может также испытать рассеяние, связанное с переходом во вторую подзону. Подчеркнем, что этот процесс, хотя и связан с изменением номера квантового уровня, является тем не менее упругим, т. е. идет с сохранением полной энергии носителя и потому может вызываться обычными упругими рассеивателями, такими как примеси или акустические фононы.

В силу сказанного выше вероятность рассеяния для носителей в первой подзоне в условиях полного вырождения может быть записана в виде

$$W_1 = W_{11} + W_{12}\Theta(\zeta - E_2), \qquad (5.8)$$

где W_{11} и W_{12} — вероятности внутри- и межподзонного рассеяния. Поскольку время релаксации, а следовательно, и подвижность обратно пропорциональны W_1, то при увеличении концентрации носителей их подвижность падает скачком при пересечении уровнем Ферми дна второй (и любой последующей) подзоны. Этот же эф-

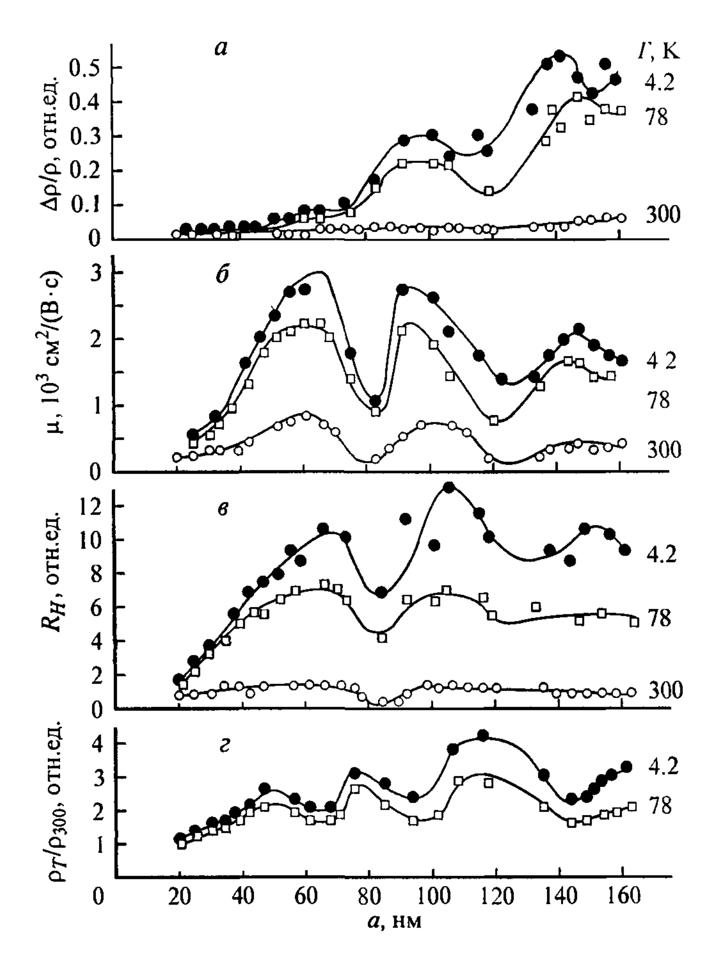


Рис. 5.3. Осцилляции магнетосопротивления (*а*), подвижности (*б*), постоянной Холла (*в*) и удельного сопротивления (*г*) в тонких пленках Bi [3].

фект должен наблюдаться, если менять не концентрацию (т. е. уровень Ферми), а положение уровней размерного квантования, например при изменении ширины квантовой ямы.

Описанный только что эффект приводит к тому, что проводимость и другие кинетические коэффициенты при фиксированной концентрации носителей в тонкой пленке являются осциллирующими функциями ее толщины. Если энергия уровней описывается формулой (1.2), то скачки проводимости должны иметь место каждый раз, когда $\zeta = \pi^2\hbar^2N^2/2ma^2$. Отсюда сразу следует, что осцилляции периодичны по толщине пленки с периодом

$$\Delta a = \frac{\pi\hbar}{\sqrt{2m\zeta}}. \qquad (5.9)$$

Подобные осцилляции являются одним из наиболее наглядных внешних проявлений размерного квантования, которое впервые наблюдалось еще в 1965 г. (рис. 5.3).

Записав закон сохранения импульса при рассеянии: $p^2/2m + E_1 = (p')^2/2m + E_2$, где p и p' — импульс в плоскости xy до и после рассеяния, видим, что межуровневое рассеяние требует большого изменения импульса носителя: $p^2 - (p')^2 = 2m(E_2 - E_1)$. Поскольку вероятность рассеяния, как правило, падает с ростом переданного при рассеянии импульса Δp,[5] то интенсивность межуровневого рассеяния обычно меньше, чем внутриуровневого. Поэтому описанные эффекты скачка подвижности и осцилляций проводимости зависят от доминирующего механизма рассеяния и для плавных рассеивающих потенциалов могут быть незаметны.

5.5. Баллистический транспорт

В двумерном электронном газе высококачественных гетероструктур с модулированным легированием длина свободного пробега электронов l может достигать нескольких десятков микрон и может оказаться сравнимой и даже превысить размеры активных областей современных полупроводниковых приборов. В этом случае электрон может пролететь через канал полевого транзистора без столкновений, как снаряд, выпущенный из пушки. Такая аналогия привела к тому, что подобные бесстолкновительные структуры часто называют баллистическими. Строго говоря, с точки зрения механики такое движение электронов не является баллистическим, поскольку оно происходит под действием постоянной ускоряющей силы. Однако этот термин стал общепризнанным.

Баллистическое движение электронов проявляется в эффекте «всплеска» скорости электронов, который оказывает большое влияние на работу приборов с малыми размерами активных областей. Этот эффект связан с конечностью времени установления стационарной энергии электронов, двигающихся в постоянном электрическом поле, и может проявляться как увеличение скорости до значений, значительно превышающих равновесные, характерные для приборов с достаточно большими размерами активных областей. На рис. 5.4 представлена зависимость мгновенного значения скорости электронов $v_{\text{мгн}}$ от расстояния L в случае постоянного электрического поля F. На малых расстояниях электроны движутся баллистически и их скорость пропорциональна времени. На некотором расстоянии от инжектирующего электрода скорость ста-

---

[5] Например, для кулоновского рассеяния дифференциальное сечение рассеяния $d\sigma \sim (\Delta p)^{-4}$.

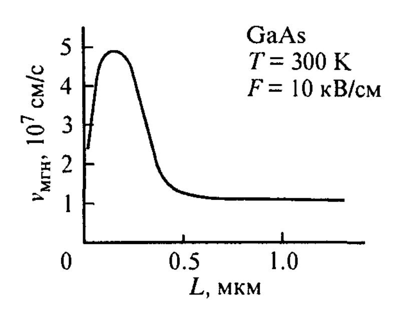


Рис. 5.4. Зависимость скорости электронов от расстояния до инжектирующего контакта для постоянного электрического поля.

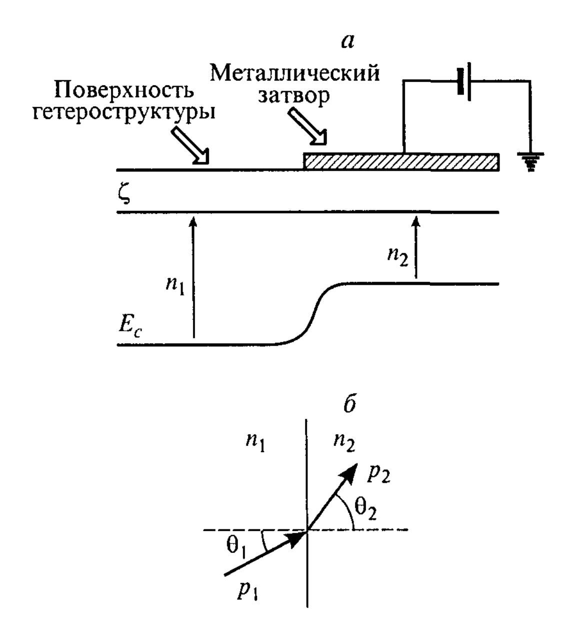


Рис. 5.5. Энергетическая диаграмма гетероструктуры с двумерным электронным газом при отрицательном потенциале затвора (*а*) и изменение импульса баллистически движущегося электрона при пересечении границы раздела двух областей двумерного электронного газа с различной концентрацией электронов (*б*).

новится максимальной, а затем уменьшается из-за рассеяния. При малых размерах активной области прибора этот эффект значительно увеличивает среднюю скорость пролета электронов через эту область, что может привести к улучшению такой важной характеристики прибора, как его быстродействие.

Эффект «всплеска» скорости проявляется в основном в приборах на основе арсенида галлия — из-за меньшей эффективной массы электронов и соответственно большей их подвижности. Этот

эффект может играть важную роль и в приборах на основе кремния, но там он выражен менее ярко — из-за большей эффективной массы электронов.

Помимо эффекта «всплеска» скорости баллистическое движение электронов в двумерном электронном газе имеет особенности, которые можно использовать для создания принципиально новых приборов наноэлектроники. В частности, представляет интерес явление преломления направления движения пучка баллистических электронов на границе раздела двух областей двумерного электронного газа с различной концентрацией электронов.

Рассмотрим гетероструктуру с двумерным электронным газом, часть которой находится под металлическим затвором (рис. 5.5, *а*). При подаче на затвор отрицательного напряжения смещения концентрация электронов под ним n_2 становится меньше концентрации n_1 в свободной области электронного газа. Уровень Ферми в обеих областях двумерного электронного газа должен остаться одним и тем же, поэтому на границе раздела областей происходит изгиб дна зоны проводимости. На электроны, пересекающие эту границу, будет действовать сила, направленная по нормали к этой границе. Для таких электронов компонента импульса, параллельная границе, сохраняется: $p_1 \sin\theta_1 = p_2 \sin\theta_2$ (рис. 5.5, *б*). В переносе электрического тока участвуют электроны с импульсом, лежащим в интервале порядка kT вблизи импульса Ферми. Связь импульса p_F с концентрацией электронов на единицу площади двумерного электронного газа может быть получена из выражения (3.12):

$$p_F = 2m(\zeta - E_1) = \hbar(\pi n_s)^{1/2}. \qquad (5.10)$$

Таким образом, для электронов, баллистически пересекающих границу раздела двух областей двумерного электронного газа, выполняется соотношение

$$\frac{\sin\theta_1}{\sin\theta_2} = \left(\frac{n_1}{n_2}\right)^{1/2}. \qquad (5.11)$$

Это выражение впервые было получено в работе [44] и является неким аналогом закона оптического преломления для баллистических электронов, пересекающих границу раздела областей двумерного электронного газа с различной концентрацией носителей заряда. Этот эффект позволяет создавать электронные призмы и линзы, а на их основе конструировать новые приборы, речь о которых пойдет в главе 8.

5.6. Вертикальный перенос в системе квантовых ям

До сих пор, рассматривая кинетические явления в двумерных электронных системах, мы ограничивались рассмотрением процессов переноса в плоскости двумерного слоя. Однако в случае квантовых ям или сверхрешеток возможно, прилагая электрическое поле по нормали к слою, вызывать ток также и в этом направлении за счет электронных переходов из ямы в яму или в широкозонные барьерные слои. Эти эффекты, называемые часто вертикальным переносом, и являются предметом изучения данного раздела.

Полевая ионизация одиночной ямы. Рассмотрим вольт-амперную характеристику (ВАХ) структуры с квантовыми ямами (см. рис. 1.9), разделенными достаточно широкими барьерами, такими, что туннелированием из ямы в яму можно пренебречь. В этом случае ток поперек слоев j может протекать только за счет выброса носителей из ямы в широкозонный барьерный слой и будет пропорционален вероятности такого выброса W. Поэтому характер ВАХ будет определяться зависимостью W от приложенного электрического поля F.

На рис. 5.6, *а* показана квантовая яма в электрическом поле. Видно, что для выброса из ямы носители должны преодолеть потенциальный барьер $E_{акт}$, равный расстоянию от уровня Ферми до правого края ямы, и потому $j \sim \exp(-E_{акт}/kT)$. Прежде чем обсуждать, к какому виду ВАХ приведет этот факт, сопоставим рассматриваемую ситуацию с эмиссией из массивного образца.

В массивном образце, как видно из рис. 5.6, *б*, величина $E_{акт}$ не зависит от поля. Вид ВАХ при этом определяется факторами, не учтенными в нашей простейшей модели: это туннельное прохождение через верхнюю часть потенциального барьера (эффект Фаулера—Нордгейма) и понижение барьера за счет сил изображения

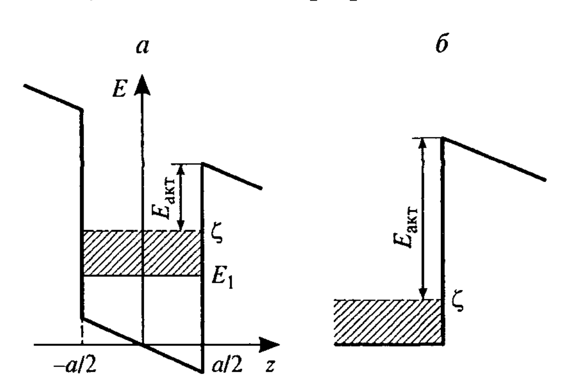


Рис. 5.6. Квантовая яма (*а*) и массивный образец (*б*) в сильном электрическом поле. Заштрихованы области, содержащие вырожденные электроны.

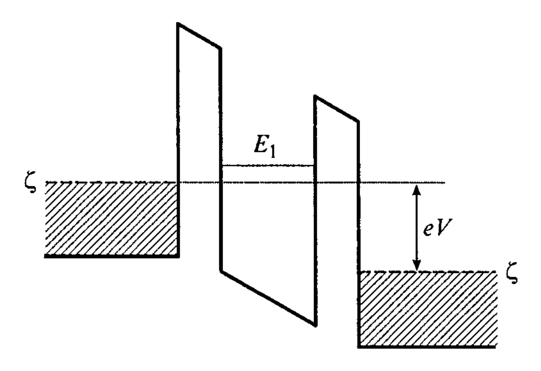


Рис. 5.7. Зонная диаграмма резонансно-туннельной структуры. Заштрихованы области, содержащие вырожденные электроны.

(эффект Шоттки) [45]. В случае квантовой ямы диэлектрические проницаемости различных слоев обычно близки и роль сил изображения не существенна. Что касается туннельных эффектов, то они могут играть роль, но при низких температурах. Если же температура достаточно высока, то ВАХ определяется зависимостью $E_{\text{акт}}(F)$, отсутствующей в массивном образце, но существующей в квантовой яме.

При не очень большой величине F ее влияние на энергию уровней в яме может быть найдено в рамках теории возмущений с возмущающим потенциалом eFz.[6] Поправка первого порядка к энергии равна $\int \Psi_1^*(z)eFz\Psi_1(z)\,dz$ [46]. Поскольку волновая функция нижнего уровня Ψ_1 — четная функция z, то эта поправка равна нулю, т. е. в линейном по полю приближении уровень не смещается. В то же время правый край ямы движется вниз с полем по закону $\Delta E - eFa/2$. Поэтому ВАХ в указанном приближении имеет вид

$$j \sim \exp\left(\frac{eFa}{2kT}\right). \tag{5.12}$$

При увеличении поля или при понижении температуры становятся существенными туннельные эффекты вблизи верхней части барьера, что приводит к постепенному снижению энергии активации в формуле (5.12).

Резонансное туннелирование. Резонансно-туннельными структурами называют совокупность полупроводниковых слоев, разделенных туннельными барьерами, где хотя бы один из слоев представляет собой квантовую яму с системой энергетических уровней (1.3). На рис. 5.7 показана зонная диаграмма простейшей резонанс-

[6] Критерием применения теории возмущений является малость безразмерного параметра $eFma^3/\hbar^2$.

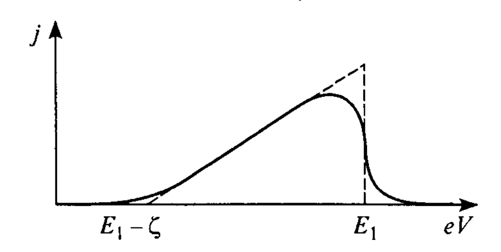


Рис. 5.8. ВАХ резонансно-туннельной структуры для простейшей модели (*штриховая кривая*) и с учетом уширения уровней (*сплошная кривая*).

но-туннельной структуры с одной квантовой ямой, двумя туннельными барьерами и двумя внешними сильно легированными массивными областями.

Если бы на месте квантовой ямы находился массивный полупроводник, то с ростом напряжения туннельные токи через оба барьера возрастали бы и ВАХ носила бы монотонный характер. Дискретность энергетического спектра в яме приводит к тому, что туннелирование через левый барьер возможно лишь тогда, когда уровень E_1 совпадает по энергии с каким-либо из заполненных состояний слева от барьера, т. е. когда падение напряжения на этом барьере V удовлетворяет условию

$$E_1 - \zeta < eV < E_1. \tag{5.13}$$

Вне этого интервала ток должен равняться нулю. Обращение тока в нуль означает, что при напряжениях, близких к правой границе интервала (5.13), ток убывает с ростом напряжения, т. е. ВАХ имеет падающий участок — участок с отрицательным дифференциальным сопротивлением. Наличие такого участка является причиной большого интереса к резонансно-туннельным структурам.

Получим формулу, описывающую конкретный вид ВАХ. Пусть L — длина массивного полупроводника, примыкающего к левому туннельному барьеру. Тогда электрон с компонентой импульса p_z пройдет эту область за время Lm/p_z и внесет вклад в ток через структуру, равный $ep_zT(p_z)/Lm$, где $T(p_z)$ — вероятность туннельного прохождения барьера для данного электрона. Для нахождения полного тока I необходимо просуммировать такие вклады по всем электронам:

$$I = \frac{2eS}{m(2\pi\hbar)^3} \int\!\!\int\!\!\int p_z T(p_z) f\left(\frac{p_z^2 + p_\parallel^2}{2m}\right) d^3p, \tag{5.14}$$

где S — площадь образца; $p_\parallel$ — компонента импульса, параллельная границе; $f(E)$ — функция распределения Ферми с химическим потенциалом ζ. Считая носители полностью вырожденными и про-

водя интегрирование по $p_{\|}$ от 0 до $\sqrt{2m\zeta - p_z^2}$, получаем для плотности тока $j = I/S$ следующее выражение:

$$j = \frac{e}{2\pi^2\hbar^3} \int_0^{\sqrt{2m\zeta}} p_z(\zeta - p_z^2/2m)T(p_z)\,dp_z\,. \tag{5.15}$$

Видно, что ВАХ определяется видом функции $T(p_z)$, т. е. туннельной прозрачностью барьера. Вообще говоря, для ее нахождения следует решать соответствующее уравнение Шредингера, но результат решения нетрудно предсказать заранее. Поскольку при туннелировании сохраняются энергия электрона и его импульс, параллельный границе, то оно может происходить лишь при энергиях, соответствующих энергии уровня в яме. В противном случае в яме не будет состояния, на которое электрон мог бы перейти. Сделанное утверждение означает, что

$$T(p_z) \sim \delta\,(p_z^2/2m + eV - E_1)\,. \tag{5.16}$$

При этом ВАХ в рассматриваемой идеализированной модели имеет треугольный вид:

$$j \sim (eV - E_1 + \zeta)\,\Theta(eV - E_1 + \zeta)\,\Theta\,(E_1 - eV), \tag{5.17}$$

показанный штриховой кривой на рис. 5.8.

В действительности энергетическое состояние E_1 в квантовой яме имеет конечное время жизни τ как за счет рассеяния, так и за счет туннельного ухода из ямы. Это приведет к конечной ширине уровня $\sim\hbar/\tau$ и к замене δ-функции в (5.16) на пик с конечными высотой и шириной. Кроме того, в реальных экспериментах температура отлична от нуля и необходимо учитывать тепловое размытие распределения Ферми. Все это приведет к тому, что особенности на ВАХ размываются и она принимает более плавный характер (сплошная кривая на рис. 5.8). Однако если температура и размытие уровней не слишком велики, тс ВАХ сохраняет падающий участок с достаточно большой амплитудой. Поскольку характерные времена туннелирования достаточно малы, этот участок сохраняется и на переменном токе вплоть до весьма больших частот сигнала. Это делает резонансно-туннельные структуры перспективными для усиления и генерации СВЧ-сигналов до субмиллиметрового диапазона.

Вольт-амперная характеристика сверхрешеток. В данном разделе будут обсуждаться ВАХ сверхрешетки с энергетическим спектром (1.10) для случая, когда внешнее электрическое поле F приложено вдоль направления периодичности (в наших обозначениях — ось z). Однако, прежде чем это делать, необходимо обсудить вопрос о том, в какой степени сам закон дисперсии (1.10)

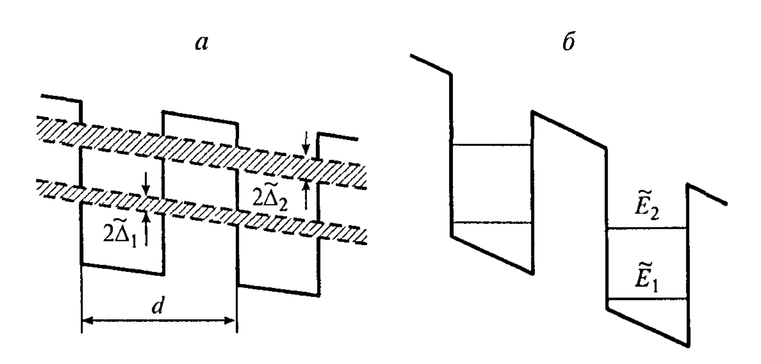


Рис. 5.9. Энергетический спектр сверхрешетки в электрическом поле. Заштрихованы разрешенные мини-зоны.

а — $eFd \ll \Delta_N$; *б* — $eFd > 2\Delta_N$.

остается справедливым в сильном электрическом поле. Вопрос этот возникает в связи с тем, что при наличии потенциала $\Phi_1 = Fz$, связанного с этим полем, суммарный потенциал, действующий на носители, складывается из потенциала периодической системы квантовых ям и потенциала Φ_1 и уже не является периодическим.

Ответ зависит от соотношения между шириной мини-зоны $2\Delta_N$, проводимость по которой мы изучаем, и разностью энергий двух соседних периодов eFd. Рассмотрим оба возможных случая.

1) $eFd \ll \Delta_N$. В этом случае разность энергий уровней в соседних ямах слишком мала, чтобы воспрепятствовать резонансному туннелированию из ямы в яму, приводящему к образованию мини-зон. Мини-зоны сохраняются, но под влиянием потенциала Φ_1 становятся наклонными (рис. 5.9, *а*). Из того же рисунка видно, что электроны, энергия которых при движении сохраняется, могут двигаться только в конечной области пространства с размером $\delta = 2\Delta_N/eF$. Подобное движение неизбежно является осциллирующим.

Физическая природа этих осцилляций заключается в следующем. Электрон, набирая энергию в электрическом поле, поднимается в верхнюю часть мини-зоны, где эффективная масса m_z для закона дисперсии (1.10), по определению равная $(\partial^2 E/\partial p_z^2)^{-1}$ [39], становится отрицательной. Меняет знак ускорение электрона, скорость его уменьшается, происходит возврат в нижнюю половину мини-зоны с положительной эффективной массой, и все повторяется снова.

Нетрудно определить частоту осцилляций. Как известно, для свободных электронов производная импульса dp/dt равна действующей на них силе. В теории твердого тела показывается, что это утверждение сохраняет силу и для произвольного закона дисперсии электронов с точностью до замены импульса на квазиимпульс [39]. Отсюда следует, что в постоянном поле F, параллельном оси

z, p_z линейно растет со временем: $p_z(t) = p_z(0) + eFt$. Подставляя это выражение в закон дисперсии (1.10), получаем, что кинетическая энергия, а следовательно, и скорость электрона осциллируют с угловой частотой

$$\Omega = \frac{eFd}{\hbar}, \tag{5.18}$$

называемой частотой Ванье—Штарка.

Мы установили, что в электрическом поле движение электрона по мини-зоне является осциллирующим и потому средний ток равен нулю и появляется лишь при учете процессов рассеяния. Ситуация здесь аналогична движению в сильном магнитном поле, которое будет рассмотрено в разделе 6.2, и по аналогии с формулой (6.6) мы можем написать выражение для проводимости вдоль оси сверхрешетки:

$$\sigma_{zz} = \frac{\sigma_0}{1 + (\Omega\tau)^2}, \tag{5.19}$$

где σ_0 — проводимость в слабых полях (при $\Omega\tau \to 0$). Поскольку плотность тока $j = \sigma_{zz} F$, то из (5.19) видно, что при $\Omega\tau \gg 1$ ВАХ имеет падающий участок вида $j \sim 1/F$.

Строго говоря, наши выводы относятся не только к сверхрешеткам, но и к любым кристаллам, поскольку мы опирались лишь на конечность ширины энергетической зоны и периодичность закона дисперсии $E(p_z)$ — общие свойства кристаллических твердых тел. Однако в обычных кристаллах наблюдать осцилляции Ванье—Штарка и связанную с ними падающую ВАХ практически невозможно. Необходимое для этого условие $\Omega\tau \gg 1$ требует столь сильных электрических полей, что их приложение к образцу вызвало бы пробой или иные необратимые изменения. В сверхрешетках, период которых на 1—2 порядка больше периода обычного кристалла, указанное условие требует значительно меньших полей и может быть без труда реализовано.

Напомним, что наши выводы получены в рамках физической картины, критерием применимости которой является условие $\hbar\Omega \ll \Delta_N$. Поскольку, с другой стороны, падающий участок на ВАХ реализуется при $\Omega\tau > 1$, наши выводы о наличии такого участка будут непротиворечивы, если $\Delta_N \tau \gg \hbar$, т. е. ширина мини-зоны превышает столкновительное уширение. Это достаточно естественное условие, поскольку только в этом случае можно вообще говорить о мини-зонном энергетическом спектре (1.10).

2) $eFd > 2\Delta_N$. В этом случае энергетическая расстройка соседних ям столь велика, что мини-зонный спектр разрушен и систему следует рассматривать как последовательность отдельных квантовых

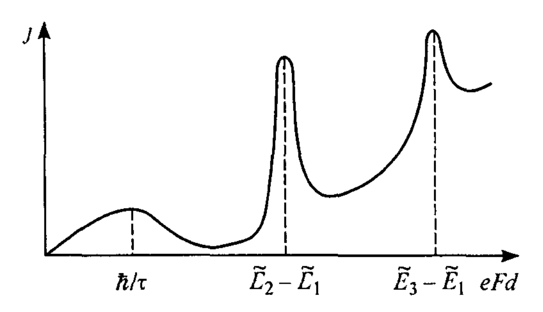


Рис. 5.10. Общий вид ВАХ сверхрешетки.

ям, сдвинутых по энергии на eFd (рис. 5.9, *б*). Как и в отдельной резонансно-туннельной структуре, рассмотренной выше, протекание тока здесь возможно лишь при совпадении энергий уровней в соседних ямах, т. е. при

$$\tilde{E}_2 - \tilde{E}_1 = eFd \tag{5.20}$$

(знак ~ указывает на то, что эти энергии сами зависят от электрического поля и потому отличаются от энергий в отсутствие поля). Если в яме более двух уровней, то возможны также резонансы при $\tilde{E}_3 - \tilde{E}_1 = eFd$ и т. д. При указанных значениях поля на ВАХ будут резкие всплески тока.

Таким образом, полная ВАХ имеет вид, изображенный на рис. 5.10. На ней имеются плавный максимум при $F = \hbar/ed\tau$ и последовательность резких резонансных пиков. На правом склоне каждого пика, естественно, также имеется падающий участок ВАХ.

Наличие падающих участков на ВАХ является важнейшим свойством сверхрешеток, позволяющим использовать их для усиления или генерации электромагнитных колебаний. Природа такой генерации особенно наглядна для падающих участков на склонах резонансных пиков. Если электрическое поле слегка больше резонансного, то первый уровень каждой ямы находится выше N-го уровня соседней на величину $\delta E = eFd - \tilde{E}_N + \tilde{E}_1$. Ситуация напоминает инверсную заселенность уровней в лазерах и допускает возможность туннельного перехода между упомянутыми выше уровнями с одновременным испусканием избыточной энергии в виде фонона с энергией δE, зависящей от электрического поля.

Для сверхрешеток, как и для резонансно-туннельных структур, можно рассчитывать на усиление и генерацию колебаний вплоть до субмиллиметрового диапазона. Именно в этой области лежат характерные собственные частоты системы (частота Ванье—Штарка и обратное время туннелирования), определяющие предельную частоту внешнего сигнала, для которой ВАХ сохраняет падающий вид.

Глава 6

СИЛЬНЫЕ МАГНИТНЫЕ ПОЛЯ И КВАНТОВЫЙ ЭФФЕКТ ХОЛЛА

Хорошо известно, что в случае сильных магнитных полей **H**, удовлетворяющих условию $\omega_c = eH/mc \gg \tau^{-1}$ и $\hbar\omega_c \gg kT$ (ω_c — циклотронная частота, τ — время релаксации), классическое описание явлений переноса неприменимо и требуется квантово-механический подход (такие магнитные поля называются квантующими) [33]. Это относится и к случаю двумерного электронного газа. Более того, как мы увидим ниже, свойства двумерных систем в квантующих полях еще более разительно, нежели в трехмерном случае, отличаются от предсказаний классической теории.

6.1. Энергетический спектр и плотность состояний двумерных систем в магнитном поле

При рассмотрении свойств двумерных систем в квантующих магнитных полях первое, что необходимо сделать, — это определить энергетический спектр системы путем решения уравнения Шредингера. Решение этой задачи достаточно хорошо известно, может быть найдено во многих учебниках, и потому мы приведем здесь лишь окончательные результаты, необходимые для нашего последующего изложения.

Прежде всего заметим, что в зависимости от ориентации магнитного поля относительно плоскости двумерного слоя могут реализоваться два случая, ответы для которых качественно различны. Если **H** лежит в плоскости слоя, то основным результатом будет увеличение энергий квантово-размерных уровней E_n за счет дополнительного ограничения движения носителей магнитным полем. Это упрощает наблюдение ряда эффектов, так как позволяет легко менять в эксперименте энергии уровней, что во всех структурах, за исключением МДП-транзисторов, в отсутствие магнитного поля является весьма непростой задачей. Происходит также некоторое увеличение эффективной массы, описывающей свободное движение в плоскости слоя. В целом можно сказать, что параллельное магнитное поле сохраняет качественный характер энергетического

спектра (1.4) и плотности состояний (3.3), хотя и меняет количественно его параметры E_n и m. Принципиально новых эффектов здесь не ожидается, и мы не будем далее обсуждать случай параллельного поля.

Ситуация, когда магнитное поле ориентировано по нормали к слою (вдоль оси z), значительно более интересна. Хорошо известно, что классические траектории частиц в плоскости, перпендикулярной магнитному полю, представляют собой окружности. В квантовой механике такое периодическое вращение отвечает системе эквидистантных дискретных уровней Ландау:

$$E = \hbar\omega_c(N + 1/2), \quad N = 0,1,2,... \tag{6.1}$$

В массивных трехмерных полупроводниках или металлах к спектру (6.1) добавляется еще непрерывный квадратичный спектр $p_z^2/2m$, связанный с движением вдоль поля, поэтому суммарная плотность состояний отлична от нуля при любой энергии, превосходящей $\hbar\omega_c/2$. Однако в двумерных системах движение по оси z квантовано, так что полный спектр представляет собой сумму (6.1) и энергий квантово-размерных уровней и, таким образом, является чисто дискретным. Удивительным является то, что чисто дискретный спектр обычно характерен для микроскопических объектов (атомы, квантовые точки и др.). Здесь же мы имеем дело с макроскопической системой.

В дальнейшем мы ограничимся рассмотрением двумерной системы с одним заполненным квантово-размерным уровнем (необходимые критерии этого приведены в разделе 3.2). Если энергию этого уровня E_1 выбрать за начало отсчета, то формула (6.1) будет описывать полный энергетический спектр системы, а не только его часть, связанную с движением в плоскости xy.

Энергетический спектр (6.1) является сильно вырожденным, т. е. каждому уровню Ландау соответствует не одно, а много электронных состояний. Физически природа вырождения достаточно понятна: в плоскости xy система однородна, что означает, что состояния, отличающиеся друг от друга лишь положением центра электронной орбиты, должны иметь одну и ту же энергию. Простые расчеты показывают (см., например, [1, 46]), что кратность вырождения (число различных состояний с одинаковой энергией) в расчете на единицу площади двумерной структуры равна

$$\eta = \frac{eH}{\pi\hbar c}. \tag{6.2}$$

Формула (6.2) имеет весьма наглядное, хотя и не строгое истолкование. Характерный размер электронной орбиты осциллятора Ландау равен так называемой магнитной длине $\lambda = \hbar c/eH$. При

этом кратность вырождения (6.2) как раз равна $(\pi\lambda^2)^{-1}$, т. е. числу электронных орбит, которое можно разместить без существенного взаимного пересечения на единице площади.

Дискретному электронному спектру (6.1) соответствует плотность состояний в виде совокупности резких пиков, в идеальном случае дельта-функций:

$$g(E) = \eta \sum_N \delta[E - \hbar\omega_c (N + 1/2)]. \qquad (6.3)$$

В реальных структурах за счет рассеяния носителей и неоднородного потенциала заряженных примесей дельта-функции будут размываться в пики с конечной шириной и высотой. Более подробно этот вопрос будет обсуждаться в разделе 6.4.

До сих пор мы не учитывали спин электрона. Наличие спина приводит к тому, что каждый уровень Ландау E_N двукратно расщепляется на состояния с энергиями $E_N \pm \beta g_e H$, где $\beta = e\hbar/2m_0 c$ — магнетон Бора, а g_e — спиновое гиромагнитное отношение (эффективный g-фактор для электронов). Эта величина, так же как и, скажем, эффективная масса носителей m, определяется деталями зонной структуры и может меняться от вещества к веществу во много раз. Используя известные значения m и g_e для наиболее распространенных полупроводников, можно показать, что величина спинового расщепления $2\beta g_e H$, как правило, заметно меньше расстояния между уровнями Ландау.

В дальнейшем мы будем учитывать спиновое расщепление уровней Ландау и, говоря о плотности состояний, брать для каждого из получившихся подуровней величину, в 2 раза меньшую, чем (6.2).

Существует простой и надежный способ экспериментального разделения эффектов орбитального и спинового квантования, основанный на использовании так называемого наклонного магнитного поля. Если **H** отклонено от нормали к двумерному слою на угол θ, то квантование Ландау (орбитальные эффекты) определяется лишь нормальной компонентой поля $H \cos\theta$, а спиновое расщепление не зависит от ориентации поля и пропорционально его полной величине H.

6.2. Кинетические явления в сильных магнитных полях

Чисто дискретный спектр двумерной электронной системы в перпендикулярном магнитном поле приводит к существенной модификации ее гальваномагнитных свойств по сравнению с трехмерным случаем. Отсутствие третьей степени свободы (свободного движения вдоль поля) приводит к тому, что вероятность упругого рассеяния носителей испытывает гигантские осцилляции с магнитным полем.

Действительно, представим себе, что напряженность магнитного поля H в точности равна $H_N = 2\pi\hbar c n_S / eN$ ($N = 1, 2,...$). Тогда при низких температурах все n_S электронов, находившиеся в двумерном электронном газе на единицу его площади, полностью заполняют N подуровней Ландау — какие-либо свободные места на них отсутствуют. Поэтому в силу принципа Паули электронных переходов внутри уровней нет. При упругом рассеянии невозможны также переходы на свободные состояния вышележащих уровней, так как это требует дополнительной энергии. Поэтому процессы рассеяния, представляющие собой переходы из одного состояния в другое под действием рассеивающего потенциала, в указанных условиях невозможны.

Чтобы понять, к каким следствиям может привести влияние магнитных полей на процессы рассеяния, рассмотрим кратко классическую картину проводимости в сильном магнитном поле. Пусть магнитное поле $\mathbf{H}$ направлено по оси z, а тянущее электрическое поле $\mathbf{F}$ — по оси x. Пусть скорость направленного движения электрона описывается вектором $\mathbf{v}$. Тогда сила, действующая на него со стороны полей, равна $e\,(\mathbf{F} + [\mathbf{v} \times \mathbf{H}]/c)$. Рассеяние носителей можно феноменологически рассматривать как эффективную силу трения, пропорциональную скорости и направленную навстречу ей: $\mathbf{F}_{mp} = -m\mathbf{v}/\tau$ (где τ — время релаксации) [39]. В постоянных полях дрейфовая скорость электрона постоянна, следовательно, сумма действующих на него сил равна нулю:

$$\frac{m\mathbf{v}}{\tau} = e\mathbf{F} + \frac{e}{c}[\mathbf{v} \times \mathbf{H}]. \tag{6.4}$$

Расписывая уравнение (6.4) по компонентам и определяя из него компоненты скорости v_x и v_y, получаем:

$$v_x = \frac{eF\tau}{m(1 + \omega_c^2\tau^2)}, \tag{6.5}$$

$$v_y = -v_x\omega_c\tau.$$

Если n_S — концентрация двумерных электронов, то величина $en_S\mathbf{v}$ представляет собой линейную (в расчете на единицу длины) плотность тока $\mathbf{j}$. Поскольку связь между векторами $\mathbf{F}$ и $\mathbf{j}$ дается тензором проводимости σ_{ik}, то из (6.5) сразу имеем выражения для диагональной (диссипативной) σ_{xx} и недиагональной (холловской) σ_{xy} компонент проводимости:

$$\sigma_{xx} = \frac{e^2 n_S \tau}{m(1 + \omega_c^2\tau^2)}, \tag{6.6}$$

$$\sigma_{xy} = \frac{e^2 n_S \omega_c \tau^2}{m(1 + \omega_c^2\tau^2)}. \tag{6.7}$$

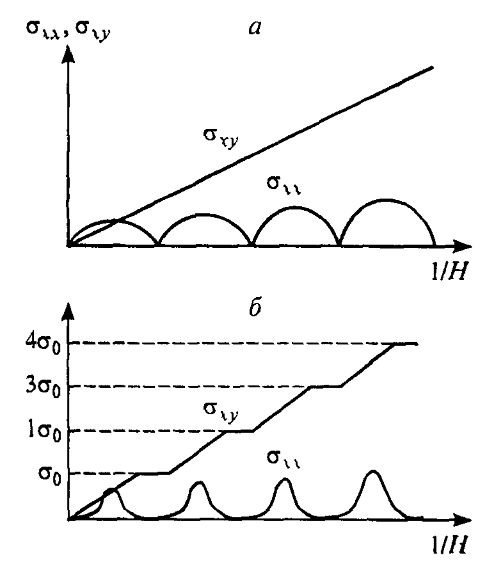


Рис. 6.1. Полевые зависимости σ_{xx} и σ_{xy} в сильных магнитных полях для идеального двумерного газа (*а*) и в реальных экспериментах (*б*) ($\sigma_0 = e^2/2\pi\hbar$).

Видно, что в сильных магнитных полях, удовлетворяющих условию $\omega_c\tau \gg 1$, диссипативная проводимость пропорциональна вероятности рассеяния τ^{-1}, а холловская проводимость не зависит от рассеяния. Физически это связано с характером движения электронов в скрещенных (взаимно перпендикулярных) полях **F** и **H**, детально описанным, например, в [47]. Оно представляет собой вращение по круговым орбитам с частотой ω_c при одновременном дрейфе центра орбиты со скоростью $\mathbf{v}_{\text{др}} = c\,[\mathbf{F} \times \mathbf{H}]/H^2$. При этом среднее смещение электрона вдоль **F** отсутствует и появляется лишь при наличии рассеяиия, сбивающего центры орбит.

Обсудим теперь зависимости σ_{xx} и σ_{xy} от магнитного поля. Как уже отмечалось, при критических значениях поля $H_N = 2\pi\hbar c n_S/eN$ рассеяние отсутствует, т. е. формально $\tau \to \infty$. Из (6.6) следует, что в этих точках σ_{xx} должна обращаться в нуль. Что касается σ_{xy}, то, согласно (6.7), она в координатах $1/H$ должна изменяться линейно, как это показано на рис. 6.1, *а*.

Для обсуждения эффектов в сильных магнитных полях очень удобно пользоваться безразмерным параметром

$$\nu = \frac{2\pi n_S \hbar c}{eH}, \tag{6.8}$$

называемым фактором заполнения. Это — число уровней Ландау (с учетом их спинового расщепления), полностью заполненных

электронами при $T = 0$. Критические поля H_N соответствуют целочисленным значениям фактора заполнения.

Весьма интересным оказывается ответ на вопрос, чему равна холловская проводимость в точках $\nu = N(N = 1, 2, 3,...)$, где исчезает диссипативная проводимость, т. е. при $H = H_N$. Подставляя выражение для H_N в (6.7), где следует также положить $\tau \to \infty$, получаем:

$$\sigma_{xy} = \frac{e^2 N}{2\pi\hbar}. \tag{6.9}$$

К этому важному выражению, не содержащему ничего, кроме мировых констант, мы вернемся в следующем разделе.

6.3. Целочисленный квантовый эффект Холла

В 1980 г. было впервые экспериментально обнаружено, что полевые зависимости компонент проводимости двумерного электронного газа кардинально отличаются от предсказаний простой теории, показанных на рис. 6.1, *а* [4, 48]. Оказалось, что диссипативная проводимость σ_{xx} действительно обращается в нуль, но не в отдельных точках, а в целых интервалах магнитных полей, причем достаточно широких (см. рис. 6.1, *б*). Еще более поразительным было поведение σ_{xy}, которая в тех же интервалах полей сохраняла постоянное значение, равное (6.9). Иными словами, холловская проводимость двумерной системы равнялась величине, не зависящей ни от параметров образца, ни от магнитного поля, ни от температуры, а определяемой только значениями фундаментальных физических констант e и $\hbar$. За открытие этого эффекта, названного квантовым эффектом Холла (далее мы будем использовать сокращение КЭХ), К. фон Клитцингу в 1985 г. была присуждена Нобелевская премия.

Заметим, что наблюдать исчезновение σ_{xx} и плато на σ_{xy} можно не только при изменении магнитного поля, но и при изменении концентрации носителей путем изменения смещения на затворе МДП-структуры.[1] Об этом говорит и формула (6.8), из которой следует, что уменьшение концентрации изменяет фактор заполнения ν таким же образом, как и увеличение магнитного поля.

Физическая модель КЭХ будет описана в следующем разделе, здесь же мы сделаем еще несколько замечаний относительно наблюдения и применения этого эффекта. Прежде всего необходимо

---

[1] Именно так был впервые обнаружен КЭХ в опытах К. фон Клитцинга.

иметь в виду, что обращение проводимости σ_{xx} в нуль не означает превращения вещества в идеальный диэлектрик, поскольку недиагональная компонента σ_{xy} при этом отлична от нуля. В условиях реальных экспериментов обычно через образец пропускают заданный ток и измеряются напряжения на контактах, т. е. измеряемой величиной является не удельная проводимость σ, а обратная величина — удельное сопротивление ρ. Вычисляя по известным формулам компоненты обратного тензора, получаем, что в областях холловских плато

$$\rho_{xx} = \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2} = 0, \quad \rho_{xy} = \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2} = \frac{2\pi\hbar}{e^2 N}. \qquad (6.10)$$

Таким образом, диагональная компонента сопротивления обращается в нуль одновременно с диагональной компонентой проводимости, а недиагональная компонента сопротивления (как и проводимости) принимает квантованные значения, равные комбинации мировых констант.

Исследования показывают, что формулы (6.9), (6.10) являются точными и не содержат поправок, т. е. с той точностью, с которой в эксперименте можно измерить сопротивление ρ_{xy}, эта величина равняется $2\pi\hbar/e^2N$. Это означает, что путем высокоточных измерений холловского сопротивления можно определить величину $e_2/\hbar$. Часто говорится [48], что КЭХ представляет собой метод прецизионного измерения постоянной тонкой структуры $e^2/\hbar c$. Это связано с тем, что значение скорости света c хорошо известно из других метрологических измерений с точностью, значительно большей, чем точность измерений e и $\hbar$ как в КЭХ, так и с помощью других методов.

Помимо измерений мировых констант, КЭХ получил и другое применение в метрологии. Величина $2\pi\hbar/e^2$, определяемая из КЭХ, имеет размерность сопротивления и численное значение ~25813 Ом. Поэтому структуру с хорошо выраженными плато КЭХ можно использовать в качестве эталона сопротивления.

6.4. Эффекты локализации и их роль в КЭХ

Рассмотрим вопрос о физической природе КЭХ, полагая для определенности, что речь идет об измерениях зависимостей σ_{xx} и σ_{xy} от концентрации электронов при фиксированном магнитном поле.

Отсутствие σ_{xx} и постоянство σ_{xy} в некотором интервале концентраций Δn_S говорят о том, что во всем этом интервале уровень химического потенциала остается зафиксированным в промежутке

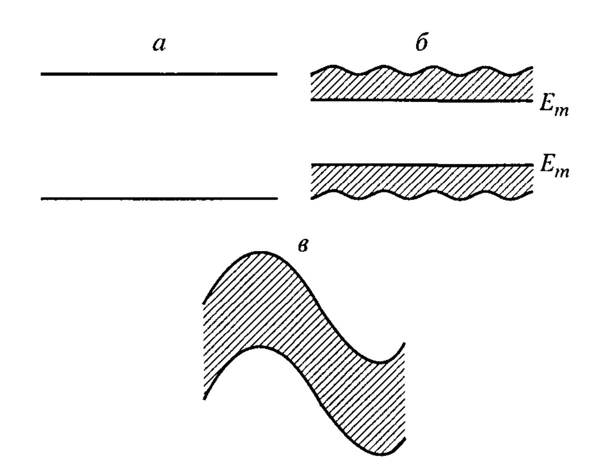


Рис 6.2. Локализация состояний в энергетической зоне с ростом потенциала неоднородностей Заштрихована область локализованных состояний.

между уровнями Ландау. Это возможно лишь в том случае, если в данном промежутке есть какие-то состояния, способные принять Δn_S электронов и не изменить при этом ток через образец. Иными словами, в образце должен существовать резервуар локализованных состояний. Современные физические представления связывают наличие таких состояний с неизбежно существующим в образце случайным потенциалом, вызванным, в частности, флуктуациями в расположении примесных ионов [49].

Основные понятия о локализации в случайном потенциале были сформулированы Андерсоном и Моттом около 40 лет тому назад. Представим себе разрешенную зону энергетических состояний (рис. 6.2, *а*). В идеальном кристалле любое состояние в зоне является делокализованным, т. е. находящиеся в нем электроны могут свободно двигаться по всему кристаллу. При наличии случайного потенциала неоднородностей $V(\mathbf{r})$ состояния вблизи края зоны становятся локализованными. Электроны этих состояний двигаются лишь в ограниченной области кристалла, не имея возможности уйти на большое расстояние, и потому не могут дать вклада в ток через образец. Вблизи центра зоны располагается область делокализованных состояний. Уровни энергии E_m, разделяющие локализованные и делокализованные состояния, называются порогами подвижности (рис. 6.2, *б*). Наконец, при большой амплитуде неоднородностей верхний и нижний пороги подвижности сливаются и все состояния становятся локализованными (рис. 6.2, *в*). Подобный переход в непроводящее состояние при увеличении амплитуды неоднородностей носит название перехода Андерсона.

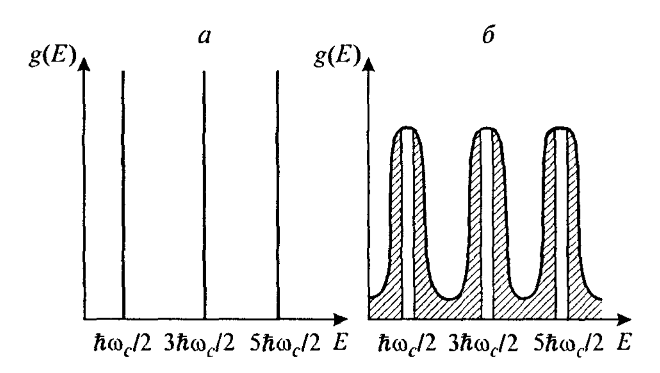


Рис. 6.3. Плотность состояний двумерной системы в перпендикулярном магнитном поле.

а — идеальный случай; *б* — при наличии потенциала неоднородностей. Заштрихована область локализованных состояний.

Изложенные представления можно перенести на случай наличия магнитного поля и применить для объяснения КЭХ. Идеальная двумерная система в перпендикулярном магнитном поле характеризуется чисто дискретным энергетическим спектром (6.1), которому отвечает плотность состояний в виде совокупности дельта-функций (6.3). Наличие случайного потенциала $V(\mathbf{r})$ приводит к тому, что разные точки пространства становятся неравноправными и энергия электрона в магнитном поле начинает зависеть от положения центра электронной орбиты. Это приведет к уширению пиков плотности состояний и появлению порогов подвижности (рис. 6.3).

Изложенных представлений достаточно для качественного объяснения КЭХ. Действительно, в интервале концентраций, когда уровень Ферми лежит в области локализованных состояний между пиками $g(E)$ (уровнями Ландау), число заполненных делокализованных состояний не меняется, а потому не меняется и σ_{xy}, ибо в проводимость дают вклад лишь делокализованные электроны. На зависимости $\sigma_{xy}(n_S)$ наблюдается плато.

Приведенные элементарные рассуждения объясняют сам факт наличия холловских плато, но не значения σ_{xy} на плато. Действительно, если делокализована лишь часть электронов, скажем, равная γn_S ($\gamma < 1$), то во всех формулах для проводимости следует, на первый взгляд, заменять n_S на γn_S, что приведет к дополнительному множителю γ в формуле (6.9). Ответ заключается в том, что потенциал неоднородностей, локализуя часть носителей, одновременно меняет и свойства делокализованных электронов. Их скорость холловского дрейфа возрастает, что компенсирует уменьшение их концентрации и приводит к сохранению фундаментальной формулы (6.9).

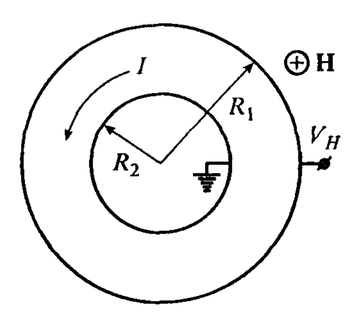


Рис. 6.4. Измерение эффекта Холла в геометрии диска Корбино.

Сделанное утверждение носит общий характер и не зависит от вида потенциала неоднородностей и геометрии системы. Чтобы доказать это, мы приведем более общий вывод формулы (6.9), не использующий результаты квантово-механических расчетов, упомянутых в разделе 6.1 и справедливых лишь для свободных электронов.

Рассмотрим образец с двумерным электронным газом, имеющий кольцевую геометрию, показанную на рис. 6.4 (диск Корбино) и содержащий некоторый потенциал неоднородностей. Пусть заполнен один уровень Ландау. Мы будем пользоваться квазиклассическим приближением и описывать электроны двумерным импульсом **p**. Движение по кольцу — периодическое и потому должно удовлетворять условиям квантования Бора—Зоммерфельда [46], которые в магнитном поле имеют вид

$$\oint \left(\mathbf{p} - \frac{e}{c}\mathbf{A} \right) d\mathbf{r} = 2\pi\hbar M \quad (M = 1, 2, 3, \ldots), \qquad (6.11)$$

где **A** — вектор-потенциал магнитного поля.

Изменим мысленно магнитное поле в отверстии кольца, не меняя его в области $r > R_1$, где находятся электроны. При этом физически наблюдаемые свойства электронов не могут меняться, поскольку они определяются величиной магнитного поля, действующего на электроны. Однако будут меняться **A** и фаза волновой функции, в которую он входит. Если полное изменение магнитного потока через отверстие будет равно $2\pi c\hbar/e$, то фазы всех волновых функций изменятся на 2π и вся электронная картина вернется в исходное состояние. На первый взгляд, ничего не изменилось. Но на самом деле при изменении **A** менялась электронная траектория. Увеличение **A** уменьшало обобщенный импульс $\mathbf{p} - e\mathbf{A}/c$, и при этом для выполнения условия квантования (6.11) должен был возрастать радиус траектории. Тот факт, что в результате описанной процедуры картина не изменилась, означает, что система уровней

приобрела исходный вид, но каждый электрон переместился на соседнюю квантованную траекторию большего радиуса, электрон с последней траектории ушел во внешний контакт при $r = R_2$, а один электрон с внутреннего контакта вошел в кольцо. Все это в целом выглядит как перемещение одного электрона с контакта $r = R_1$ в контакт $r = R_2$.

Рассмотрим баланс энергии при описанном выше действии. Поскольку разность потенциалов между указанными контактами равна холловскому напряжению V_H, при этом совершается работа eV_H. Взглянем на проблему с другой стороны. При изменении магнитного поля в отверстии всюду, в том числе и в плоскости, содержащей электроны, возникало индукционное электрическое поле. Согласно законам электрической индукции, при изменении на $\Delta\varphi$ магнитного потока через контур с током I энергия системы меняется на $\Delta E = I\Delta\varphi/c$. В нашем случае $\Delta\varphi = 2\pi c\hbar/e$ и $\Delta E = 2\pi\hbar I/e$. Поскольку уровень Ферми лежит в области локализованных состояний и $\sigma_{xx} = 0$, то диссипативных токовых потерь индукционное поле не вызывает. Поэтому из баланса энергии следует, что $\Delta E = eV_H$, т. е.

$$\frac{I}{V_H} = \sigma_{xy} = \frac{e^2}{2\pi\hbar}. \tag{6.12}$$

Аналогичные рассуждения для системы с N заполненными уровнями Ландау дали бы дополнительный множитель N в формуле (6.12). Высказанные аргументы могут служить доказательством универсального характера формулы КЭХ (6.9).

6.5. Дробный КЭХ

Через несколько лет после экспериментального открытия и теоретического объяснения описанного выше целочисленного КЭХ при исследовании образцов с очень высокой подвижностью двумерных носителей ($> 10^6$ см^2/(В · с)) при сверхнизких температурах (< 1 К) было обнаружено новое неожиданное явление, названное дробным КЭХ. Это дополнительные плато на полевых зависимостях σ_{xy} (или ρ_{xy}), описываемые формулой (6.9), где вместо целого числа N стоит дробь вида k/n, причем n — нечетное число, а k — любое целое. С формальной точки зрения так мог бы выглядеть обычный КЭХ, если бы заряд электронов имел дробную величину. При значениях полей, отвечающих плато σ_{xy}, диссипативная проводимость σ_{xx} если и не обращается в нуль, то во всяком случае имеет резкий минимум. К настоящему времени подобный дробный КЭХ наблюдался для знаменателей n вплоть до 13, так что при

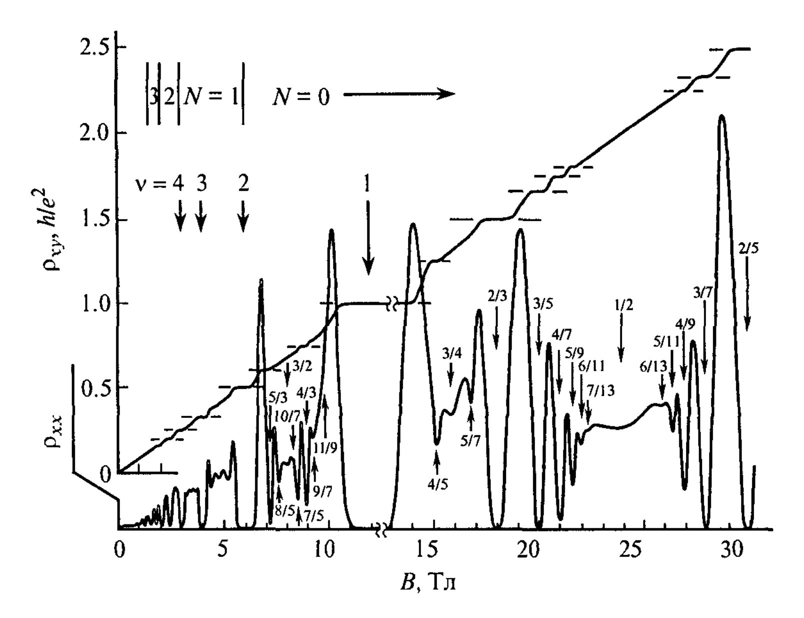


Рис. 6.5. Дробный КЭХ в двумерном электронном газе при низкой температуре и сильных магнитных полях (масштаб кривой $\rho_{xx}(B)$ в сильных полях увеличен в 2.5 раза) [50].

факторах заполнения $\nu < 1$ полевые зависимости гальваномагнитных коэффициентов могут быть насыщены особенностями, как это показано на рис. 6.5.

Несмотря на внешнее сходство с целочисленным КЭХ, указанный дробный эффект имеет совсем иную природу. Он не может быть объяснен в рамках простого одноэлектронного подхода, использованного в предыдущем разделе, а требует учета электрон-электронных взаимодействий. Эти взаимодействия в сильном магнитном поле приводят к тому, что электронный газ превращается в особую сильно коррелированную жидкость [48]. В такой жидкости можно выделить связанные состояния, содержащие 3, 5, 7 и т. д. электронов (нечетность этих чисел связана с принципом Паули, который формально может быть представлен как требование нечетности электронной волновой функции по отношению к перестановке любых двух электронов). Можно показать, что возбуждения в таких состояниях будут иметь эффективный заряд соответственно 1/3, 1/5 и т. д.[2] Формально подобные возбуждения

[2] Это не так удивительно, как может показаться на первый взгляд. Если в многочастичной системе, включающей свободные электроны и ионы решетки, возбуждения (электроны проводимости и дырки в валентной зоне) имеют эффективную массу, отличную от массы свободного электрона, то почему в другой многочастичной системе заряд возбуждений не может отличаться от заряда электронов?

часто рассматривают как так называемые композитные фермионы, представляющие собой обычный фермион (электрон), связанный с четным числом квантов магнитного потока. В отличие от исходных сильно взаимодействующих электронов композитные фермионы уже могут считаться не взаимодействующими друг с другом частицами. Применяя аргументы предыдущего раздела не к обычному электронному газу, а к газу таких возбуждений, получим теоретическую модель, описывающую дробный КЭХ.

Глава 7

СВОЙСТВА КВАНТОВЫХ НИТЕЙ И ТОЧЕК

7.1. Баллистическая проводимость нитей

Характер проводимости квантовых нитей существенно зависит от их длины. Если длина нити L существенно превосходит длину свободного пробега носителей l, то удельная проводимость нити в расчете на единицу длины дается обычным классическим выражением $\sigma = ne^2l/mv$, где v — тепловая скорость носителей. Различие с массивным образцом будет лишь в том, что рассеяние носителей в нити и в массивном образце может быть весьма разным, в связи с чем длина пробега l в нити дается совсем иными выражениями и требует специального анализа для своего нахождения.

Значительно более интересна и необычна ситуация $L < l$ [6]. При этом электрон, вылетев из одного контакта, долетает до другого контакта без столкновений, как снаряд, выпущенный из пушки. Такая аналогия привела к тому, что рассматриваемые бесстолкновительные структуры часто называют баллистическими.

Пусть имеется баллистическая одномерная структура, снабженная металлическими контактами, между которыми мы приложили напряжение V. Контакты можно рассматривать как электронные резервуары, характеризуемые химическими потенциалами ζ_1 и ζ_2, причем $\zeta_1 - \zeta_2 = eV$. Для простоты будем считать температуру достаточно низкой, так что электроны в резервуарах полностью вырождены. В области энергий $E < \zeta_2$ состояния в левом и правом контактах полностью заполнены, так что электроны из этих состояний не могут создавать ток в цепи. Такой ток связан исключительно с электронами из энергетического интервала $\zeta_2 < E < \zeta_1$, где в левом контакте есть электроны, вылетающие в нить, а состояния правого контакта пусты и способны эти электроны принять. Вычислим величину возникающего тока.

Если электрон имеет импульс p_x и, следовательно, скорость p_x/m, то его вклад в ток равен ep_x/mL. Для получения полного тока I

нам необходимо сложить такие вклады от всех электронов с энергиями $\zeta_2 < E < \zeta_1$ во всех подзонах:

$$I = \frac{e}{mL} \sum_{m,n,E} p_x, \quad \zeta_2 < E_{mn} + \frac{p_x^2}{2m} < \zeta_1 . \qquad (7.1)$$

Если разность $\zeta_2 - \zeta_1$ мала, то интересующие нас электроны существуют только в подзонах с $E_{mn} < \zeta_2$, причем импульсы их лежат в интервале длиной $\Delta p = eVm/p_x$ вблизи $[2m(\zeta_2 — E_{mn})]^{1/2}$. Из сказанного в главе 3 следует, что в интервале Δp с учетом спинового вырождения есть $2\Delta pL/(2\pi)$ различных электронных состояний, и окончательно из (7.1) имеем $I = e^2 VN/\pi$, где N — число уровней (подзон), лежащих ниже уровня химического потенциала, т. е. содержащих электроны.

Итак, наши выкладки привели к тому, что проводимость баллистической квантовой нити

$$\sigma = \frac{e^2}{\pi\hbar} N. \qquad (7.2)$$

Полученная формула носит общий характер и не зависит ни от характеристик нити (за исключением числа заполненных уровней), ни от условий измерений. Как и формула для КЭХ (6.9), она не содержит ничего, кроме мировых констант.

Необходимо обратить внимание еще на один важный вопрос. Наличие конечной проводимости у системы означает, что при приложении к ней напряжения V в системе протекает ток и происходит выделение энергии, равное $\sigma V^2/L$ в единицу времени. Это аналог эффекта Джоуля—Ленца в обычном проводнике. Физика джоулевых потерь хорошо известна: электроны, разгоняющиеся в электрическом поле, отдают энергию кристаллической решетке за счет столкновений. Но в баллистической нити электроны не испытывают столкновений! Откуда же берутся тепловые потери? Они происходят не в самой нити, а в контактах, причем в обоих контактах поровну. Известно, что в системе вырожденных электронов весь токоперенос осуществляется электронами на уровне Ферми. Иными словами, все электроны, поступающие в левый контакт из внешней цепи, имеют энергию ζ_1. Уходят же из контакта в нить электроны из интервала энергий $\zeta_2 < E < \zeta_1$, т. е. со средней энергией $(\zeta_2 + \zeta_1)/2$. Таким образом, если мы считаем, что распределение электронов в левом контакте равновесно и не меняется со временем, то каждый электрон, приходящий из внешней цепи, должен за счет рассеяния в контакте отдать кристаллической решетке энергию, в среднем равную $(\zeta_1 - \zeta_2)/2$. Аналогичная ситуация и в правом контакте. В него из нити поступают электроны с энергиями от ζ_2 до ζ_1. Приходя в равновесие, они должны «остыть» до значения ζ_2 и тем самым отдать энергию, также в среднем равную $(\zeta_1 - \zeta_2)/2$.

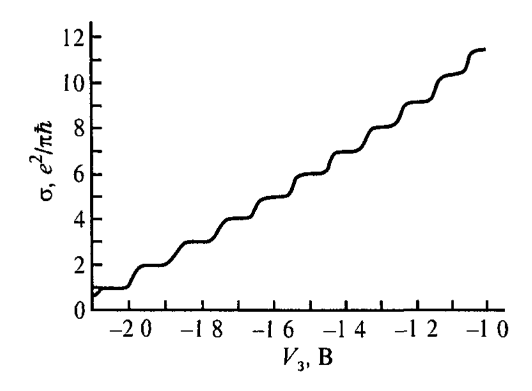


Рис 7.1. Экспериментальная зависимость проводимости квантовой нити (точечного контакта) от напряжения на затворе, определяющего концентрацию носителей [51].

Интересно проследить, как зависит проводимость баллистической нити от концентрации электронов в ней. В структурах с затвором Шоттки, описанных в разделе 1.4, концентрацию можно менять с помощью напряжения, прикладываемого к затвору. По мере уменьшения отрицательного потенциала на затворе ширина квантовой нити и концентрация носителей в ней возрастают. При этом возрастет число заполненных квантово-размерных подзон, как за счет роста концентрации, т. е. уровня Ферми, так и за счет уменьшения энергетического расстояния между квантовыми уровнями. Если температура низка и электроны в структуре полностью вырождены, то каждый раз, когда очередной уровень E_{mn} пересекает уровень Ферми, число N в формуле (7.2) меняется на единицу и проводимость нити возрастает на $e^2/\pi\hbar$. В целом зависимость σ от напряжения на затворе будет иметь вид ступенчатой функции, показанной на рис. 7.1.

Если нить не является достаточно короткой и высококачественной, то электрон на пути от контакта до контакта может испытать рассеяние на примеси или ином дефекте нити. Такое рассеяние является упругим, т. е. происходит без изменения энергии. Если электрон остается на том же квантовом уровне E_{mn}, то упругое рассеяние можно осуществить лишь одним образом: сменив импульс p_x на $-p_x$, т. е. повернув строго назад.[1] Естественно, что ток при этом уменьшается. Если для электрона из подзоны mn вероятность такого отражения равна R_{mn}, то вместо формулы (7.2) мы будем иметь:

$$\sigma = \frac{e^2}{\pi\hbar}\sum_{m,n}(1 - R_{mn}). \qquad (7.3)$$

---

[1] В дву- и трехмерном случаях возможности для упругого рассеяния значительно больше· вектор импульса, не меняясь по величине, может повернуться на любой угол по отношению к своему начальному направлению.

За счет зависимости коэффициентов отражения R_{mn} от энергии электронов, а также за счет теплового размытия функции распределения носителей ступеньки, даваемые формулой (7.3), могут быть несколько размыты, однако в реальных высококачественных структурах они могут наблюдаться очень отчетливо.

Экспериментальные исследования описанного квантования проводимости в коротких (баллистических) нитях часто проводятся на структурах, представляющих собой, строго говоря, не нить как таковую, а так называемый точечный контакт, т. е. узкую перемычку, соединяющую собой два участка двумерного электронного газа достаточно большой площади. Формально это как бы нить, имеющая длину, сравнимую с ее шириной, причем последняя имеет достаточно малую величину. Квантование проводимости должно наблюдаться и в таких структурах (это неудивительно, поскольку окончательная формула (7.2) не содержит никаких конкретных параметров, описывающих размеры и форму нити). В частности, экспериментальные зависимости, показанные на рис. 7.1, получены именно на таком квантовом микроконтакте.

7.2. Связь КЭХ с квантованием проводимости в нитях

Формула (7.2), предсказывающая квантование проводимости в баллистических квантовых нитях, совершенно неожиданно имеет такой же вид, как и формула (6.9) из теории КЭХ. Оказывается, такое совпадение не случайно. Несмотря на то что в отсутствие магнитного поля одномерные системы, какими являются квантовые нити, принципиально отличаются от двумерных электронных систем, последние в сильном магнитном поле приобретают черты, свойственные квантовым нитям.

Чтобы понять, в чем здесь дело, надо рассмотреть общие черты динамики двумерных электронов в магнитном поле **H**, перпендикулярном их плоскости. Как отмечалось в разделе 6.1, электроны в **H** движутся по круговым орбитам с частотой ω_c, что с точки зрения квантовой механики отвечает энергетическому спектру в виде эквидистантных уровней Ландау. Это, однако, относится лишь к электронам в центральной части образца, вдали от его границ. Совсем иначе ведут себя электроны, находящиеся вблизи границы образца. Как видно из рис. 7.2, за счет многократных отражений от границы они получают возможность поступательного движения вдоль края. Это движение может быть охарактеризовано некоторым средним значением импульса p_x, параллельного границе, и соответствующей кинетической энергией E. Таким образом, динамика приграничных двумерных электронов в сильном магнитном поле напоминает их динамику в квантовых нитях, где электроны

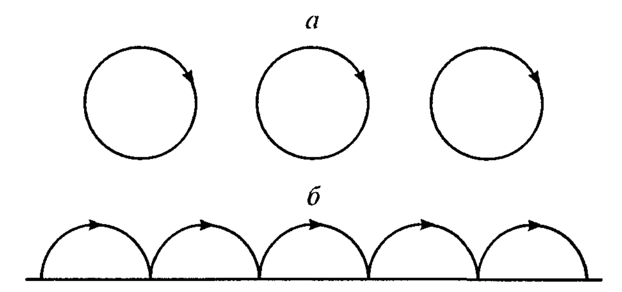


Рис. 7.2. Траектории двумерных электронов в перпендикулярном магнитном поле в глубине образца (*а*) и вблизи границы (*б*).

могут двигаться в одном направлении и резко ограничены в своем движении в двух других.

И еще одно важное замечание. В квантовой нити без магнитного поля ток, параллельный оси нити, создается электрическим полем, имеющим то же самое направление. В магнитном поле сила, действующая на электрон со стороны электрического поля (сила Лоренца), перпендикулярна как электрическому, так и магнитному полю. Поэтому формула типа (7.2) в магнитном поле будет связывать между собой x-компоненту электрического поля с y-компонентой тока, и наоборот. Итак, мы можем окончательно утверждать, что в двумерном электронном газе в сильном магнитном поле недиагональная холловская компонента проводимости дается выражением (7.2), что составляет основное содержание КЭХ и может рассматриваться как его альтернативный качественный вывод, использующий иную физическую модель, нежели в главе 6.

Важным отличием КЭХ от проводимости квантовой нити является тот факт, что вдоль данной границы электроны могут двигаться лишь в одну сторону. Движение в другую сторону осуществляется вдоль противоположной границы образца. При этом становится невозможным упругое отражение электронов, т. е. смена знака импульса за счет упругого рассеяния на примесях или дефектах. Такой процесс требовал бы перескока электрона на макроскопическое расстояние от одной границы до другой, что невозможно при рассеянии. Поэтому в отличие от нитей, где рассеяние носителей назад способно привести к значительным отклонениям проводимости от универсального значения (7.2) (см. формулу (7.3)), в условиях КЭХ данный результат является точным.

7.3. Кулоновская блокада

Помимо большой совокупности рассмотренных в предыдущих главах интересных физических свойств, которыми обладают одиночные квантовые ямы, нити и точки, существует еще одна группа

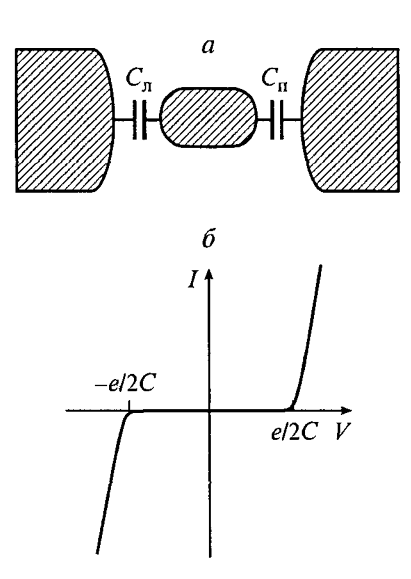


Рис. 7.3. Схематическое изображение структуры для наблюдения явления кулоновской блокады (*а*) и ее вольт-амперная характеристика (*б*).

явлений, связанных с прохождением тока через структуры, отделенные от массивных внешних электродов туннельно-прозрачными барьерами. На рис. 7.3, *а* представлено схематическое изображение такой структуры с квантовой точкой в качестве центрального электрода. В главе 5 мы обсуждали эффекты резонансного туннелирования через квантовую яму. В данном разделе мы рассмотрим еще одно, принципиально новое явление, возникающее при туннелировании носителей через квантовую точку.

Это явление, называемое кулоновской блокадой, опирается на тот факт, что заряд не может дробиться бесконечно и переносится через барьер дискретными порциями, кратными заряду электрона [52, 53]. Если размер, а следовательно, и емкость квантовой точки $C = C_л + C_п$ ($C_л$ и $C_п$ — емкости левого и правого туннельных переходов) достаточно малы, то перенос одного электрона в точку или из нее меняет электростатическую энергию на величину $e^2/2C$. Указанный энергетический барьер должен быть преодолен в ходе токопереноса. Это означает, что при малых напряжениях, приложенных к туннельной структуре, ток будет отсутствовать и возникнет лишь тогда, когда приложенное напряжение V превзойдет по абсолютной величине $e/2C$ (рис 7.3, *б*). Иными словами, вольт-амперная характеристика туннельного контакта при малых V будет существенно нелинейна.

Для наблюдения эффектов кулоновской блокады требуются достаточно малые размеры квантовых точек и низкие температуры. Действительно, упомянутый электростатический потенциальный

барьер будет существенно изменять свойства структуры, если его величина заметно превосходит тепловую энергию носителей:

$$\frac{e^2}{2C} \gg kT. \tag{7.4}$$

Например, для структуры, где квантовая точка имеет характерный диаметр 0.1 мкм и отделена от контакта диэлектрическим слоем толщиной 20 нм, емкость $C \approx 5 \cdot 10^{-17}$ Ф. При этом условие (7.4) соответствует требованию $T \ll 30$ K. Для наблюдения эффекта кулоновской блокады при более высоких температурах необходимо уменьшать размеры квантовой точки, а значит, и ее электрическую емкость.

Следует подчеркнуть, что, хотя эффекты кулоновской блокады для своей реализации также требуют малых размеров точки a, условие (7.4), налагаемое на a, существенно иное, нежели условия для наблюдения квантовых размерных эффектов, сформулированные в главе 1. Поэтому эффекты кулоновской блокады и размерного квантования являются независимыми. В частности, кулоновская блокада наблюдается в таких, например, структурах, как островковые пленки металла, где в силу достаточно большой эффективной массы электронов и высокой энергии Ферми размерное квантование практически не наблюдаемо.

На основе явления кулоновской блокады может быть изготовлен принципиально новый прибор — одноэлектронный транзистор, речь о котором пойдет в следующей главе.

Глава 8

ПРИМЕНЕНИЕ КВАНТОВО-РАЗМЕРНЫХ СТРУКТУР В ПРИБОРАХ МИКРО- И НАНОЭЛЕКТРОНИКИ

Специфические электронные свойства низкоразмерных систем открывают широкую перспективу для создания на их основе новых типов полупроводниковых приборов. Некоторые возможности практического применения структур, ВАХ которых содержит участки отрицательного дифференциального сопротивления, уже обсуждались в главе 5. Здесь будут рассмотрены другие аспекты приборного применения квантово-размерных структур.

8.1. Лазеры с квантовыми ямами и точками

Самым распространенным типом полупроводникового лазера является лазер на двойной гетероструктуре, где активная область представляет собой тонкий слой узкозонного полупроводника между двумя широкозонными [54]. Двойные гетероструктуры предложены в 1963 г. Ж. И. Алфёровым. В 2000 г. «за развитие полупроводниковых гетероструктур, используемых в высокоскоростной электронике и оптоэлектронике», Ж. И. Алфёров и Г. Крёмер (США) награждены Нобелевской премией по физике. При достаточно малой толщине активной области a она начинает вести себя как квантовая яма и квантование энергетического спектра в ней существенно меняет свойства лазеров [13].

Основное влияние на свойства лазеров оказывает изменение плотности состояний, происходящее под влиянием размерного квантования (см. раздел 3.1). Если в массивном полупроводнике в непосредственной близости от края зоны эта величина мала (см. рис. 3.1), то в квантово-размерной системе она не убывает вблизи края, оставаясь равной $m/\pi\hbar^2$. Благодаря этому факту условия создания инверсной населенности в двумерных системах оказываются более благоприятными, чем в трехмерных. Это привело к конкретным практическим результатам. Создание лазеров с квантово-размерной активной областью позволило получить непрерывную генерацию при комнатной температуре и в дальнейшем снизить пороговый ток инжекционного лазера до рекордно низких значе-

ний, составляющих к настоящему моменту величину порядка 50 А/см^2. Развитию полупроводниковых инжекционных лазеров посвящен обзор Ж. И. Алферова «История и будущее полупроводниковых гетероструктур» (ФТП. 1998. Т. 32, № 1. С. 3—18).

Благодаря иной энергетической зависимости плотности состояний меняется не только величина порогового тока, но и его температурная зависимость. Она становится более слабой, в силу чего непрерывную генерацию удается получить не только при комнатной температуре, но и при температурах на много десятков градусов выше.

Другой важной особенностью лазеров на квантовых ямах является возможность их частотной перестройки. Минимальная энергия излучаемых световых квантов равна $\hbar\omega = E_g + E_1^e + E_1^h$. Она меняется при изменении a, т. е. путем изменения ширины квантовой ямы можно осуществлять перестройку частоты генерации, сдвигая ее в коротковолновую сторону по сравнению с лазерами с широкой (классической) активной областью.

В квантовых точках энергетический спектр меняется еще более радикально, чем в квантовых ямах. Плотность состояний имеет δ-образный вид, и в результате отсутствуют состояния, которые не принимают участия в усилении оптического излучения, но содержат электроны. Это уменьшает потери энергии и как следствие уменьшает пороговый ток. Лазеры могут содержать одну или (для увеличения оптического усиления) несколько плоскостей, заполненных квантовыми точками. Используются также вертикально связанные точки (см. раздел 2.5) [55]. Согласно теоретическим оценкам, диодные лазеры с активной средой из квантовых точек должны обладать значительно лучшими свойствами по сравнению с лазерами на квантовых ямах, а именно: существенно бóльшим коэффициентом усиления, меньшей пороговой плотностью тока, полной невосприимчивостью к температуре решетки, лучшими динамическими характеристиками и бóльшими возможностями для контроля за энергией кванта излучения [16]. Показано, что в лазерах на основе вертикально связанных квантовых точек пороговый ток при комнатной температуре может быть снижен до 15 А/см^2 [56].

8.2. Оптические модуляторы

Структуры с квантовыми ямами могут использоваться не только для генерации, но и для модуляции светового излучения из области межзонного оптического поглощения. Физической основой эффекта модуляции являются два следующих экспериментальных факта:

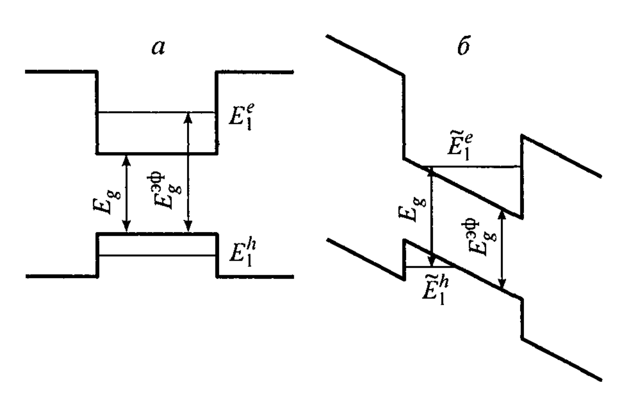


Рис. 8.1. Энергетический спектр квантовой ямы в отсутствие (*а*) и при наличии (*б*) приложенного электрического поля.

а) в структурах с квантовыми ямами существуют резкие линии экситонного поглощения, имеющие значительно бóльшую интенсивность, чем в однородных полупроводниках, и наблюдаемые не только при низких, но и при комнатных температурах;

б) спектральное положение линий заметно сдвигается при приложении сильного электрического поля перпендикулярно слоям гетероструктуры.

Первое обстоятельство уже отмечалось в разделе 3.4, а второе нуждается в дополнительном комментарии.

В однородных полупроводниках электрическое поле разрушает экситоны, растаскивая электрон и дырку в разные стороны. Наличие же квантовой ямы позволяет без вреда для экситона приложить в поперечном направлении достаточно сильное поле, поскольку стенки ямы не дают электрону и дырке возможности разойтись. Поэтому экситонная линия поглощения в поле не исчезает, но положение ее достаточно сильно сдвигается. Это обусловлено двумя факторами. Во-первых, поле искажает форму самой квантовой ямы, превращая ее из прямоугольной в трапециевидную (рис. 8.1). При этом меняются энергии квантовых уровней как в зоне проводимости $\tilde{E}_N^e$, так и в валентной зоне $\tilde{E}_N^h$, а следовательно, и эффективная ширина запрещенной зоны $E_g^{эф} = E_g + \tilde{E}_1^e + \tilde{E}_1^h$. Во-вторых, энергия связи экситона E_{ex} хотя и не обращается в нуль, как только что было отмечено, но тем не менее зависит от приложенного поля. В результате спектральное положение экситонной линии $\hbar\omega_{ex} = E_g^{эф} - E_{ex}$ оказывается сильно зависящим от напряженности электрического поля F.[1]

Поскольку экситонная линия в спектре поглощения является весьма резкой, то коэффициент поглощения α для света с частотой

---

[1] Поскольку в двумерных системах, как правило, $E_{ex} \ll E_1^e$, то зависимость $\hbar\omega(F)$ определяется в основном первым фактором (сдвигом энергии квантово-размерных уровней).

вблизи этой линии очень сильно меняется при изменении положения линии. Это делает весьма эффективной модуляцию света с помощью приложенного электрического поля. Ожидается, что подобные модуляторы найдут применение в системах оптической передачи информации.

8.3. Фоточувствительные *nipi*-структуры

Краткое представление о дельта-легировании было дано в разделе 1.3. Наряду с отдельными δ-слоями в ряде случаев изготавливаются и исследуются структуры, содержащие много подобных слоев. Практический интерес представляют структуры, содержащие не однотипные слои, а так называемые *nipi*-сверхрешетки. Это периодическая последовательность чередующихся δ-слоев *n*- и *p*-типа проводимости. Если концентрация примесей в *n*- и *p*-слоях одинакова, то система в равновесии может вообще не содержать свободных носителей. Возможности практического применения таких сверхрешеточных структур в качестве фоточувствительных элементов обусловлены двумя их основными отличиями от однородных полупроводников: длинноволновым сдвигом края фоточувствительности и очень большим временем жизни фотовозбужденных носителей. Оба эти обстоятельства проиллюстрированы на рис. 8.2.

Как видно из рисунка, минимальное энергетическое расстояние между низшим состоянием зоны проводимости в *nipi*-структуре и высшим состоянием в валентной зоне $E_g^{эф}$ значительно меньше ширины запрещенной зоны полупроводника. Это делает возможным оптические переходы между такими состояниями, т. е. межзонное поглощение при энергиях кванта $h\nu$ меньше ширины запрещенной зоны E_g. Поглощение будет существовать в меру перекрытия волновых функций электронов и дырок. Амплитуда волновой функции электрона в месте нахождения дырок будет пропорциональна вероятности туннелирования W под треугольный барьер высотой $E_g - E_g^{эф}$ на расстояние $(E_g - E_g^{эф})/eF$, где F — напряженность внутреннего электрического поля в сверхрешетке. Для сверхрешетки с равными концентрациями доноров и акцепторов в соседних слоях N_i $F = 4\pi e N_i/\varepsilon$, где ε — диэлектрическая проницаемость полупроводника.[2]

Упомянутая вероятность туннелирования равна:

$$W = \exp\left[-\frac{4\sqrt{2m}(E_g - E_g^{эф})^{3/2}}{3e\hbar F}\right].$$

---

[2] Если плотности доноров и акцепторов в *n*- и *p*-слоях не равны, в роли N_i выступает меньшая из этих величин.

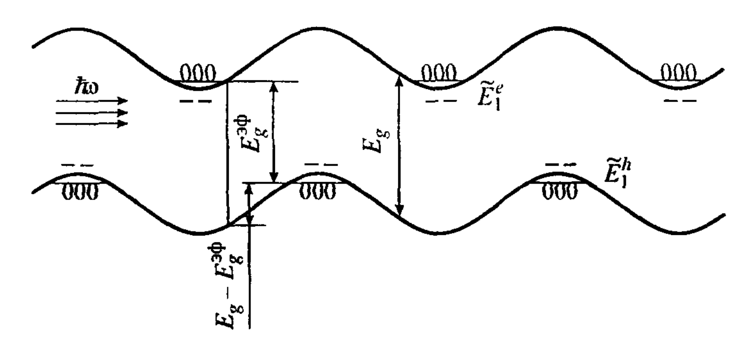


Рис. 8.2. Энергетическая диаграмма *nipi*-сверхрешетки при освещении.

Поэтому в сильно легированных *nipi*-сверхрешетках с большим внутренним полем возможно поглощение фотонов с энергией, существенно меньшей E_g. Это, по сути дела, эффект Франца—Келдыша во внутреннем электрическом поле сверхрешетки.

Большое значение времени жизни связано с пространственным разделением неравновесных носителей. Неравновесные электроны располагаются вблизи минимумов зоны проводимости, а дырки — вблизи максимумов валентной зоны. Как можно видеть из рис. 8.2, для того чтобы оказаться в одной точке пространства и рекомбинировать, им надо преодолеть потенциальный барьер $E_g - E_g^{эф}$, поэтому время жизни

$$\tau = \tau_0 \exp\left(\frac{E_g - E_g^{эф}}{kT}\right), \qquad (8.1)$$

где τ_0 — время жизни однородного полупроводника. Подобное возрастание времени жизни является общим свойством неоднородных полупроводников [57].

Большое время жизни в значительной степени компенсирует туннельную малость коэффициента поглощения, в результате чего *nipi*-сверхрешетки имеют заметную фоточувствительность в значительной спектральной области за краем собственного поглощения и потому могут использоваться в качестве фотоприемников в области бóльших длин волн, нежели однородные полупроводники. Недостатком таких приемников будет значительная инерционность, определяемая большим временем жизни (8.1).

8.4. Фотоприемники на квантовых ямах

Процессы оптической ионизации квантовых ям, описанные в разделе 4.3, могут использоваться для создания новых типов приемников инфракрасного излучения. Принцип приемника весьма прост: выброс носителей в зону проводимости широкозонного по-

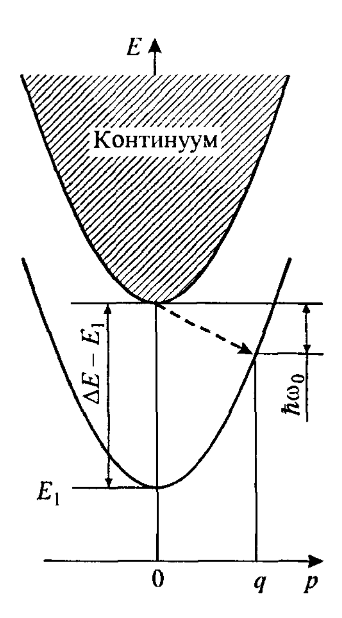


Рис. 8.3. Процесс захвата неравновесного электрона в квантовую яму с испусканием оптического фонона.

лупроводника (потенциального барьера) увеличивает проводимость в направлении, перпендикулярном слоям гетероструктуры.

По своему действию такой приемник напоминает примесный фоторезистор, где в роли центров выступают квантовые ямы. Поэтому в качестве времени жизни неравновесных носителей — важнейшей характеристики фоточувствительного материала — выступает характерное время захвата в квантовую яму τ_q. По сравнению с обычным временем жизни, связанным с захватом на рекомбинационные центры, τ_q обладает двумя важными отличиями.

Во-первых, τ_q значительно (на несколько порядков) меньше времени захвата на центры. Причина в том, что акт захвата связан с необходимостью передачи решетке от носителя достаточно большой энергии, равной энергии связи центра или же величине ΔE при захвате в квантовую яму. Наиболее эффективный механизм передачи энергии — это испускание оптических фотонов с энергией $\hbar\omega_0$. Однако энергия связи центров отнюдь не совпадает с $\hbar\omega_0$, и потому такой процесс невозможен. Электрон должен отдавать энергию в ходе значительно более медленного каскадного процесса испускания многих акустических фононов [58]. В случае квантовой ямы наличие непрерывного спектра движения в плоскости ямы (1.4) существенно меняет ситуацию. Становится возможным переход на связанное состояние в яме при испускании оптического фонона с одновременной передачей оставшейся избыточ-

ной энергии в движение в плоскости ямы (рис. 8.3). Если исходный электрон имел энергию, близкую к краю зоны в широкозонном материале, то из рис. 8.3 видно, что испускаемый фонон должен иметь достаточно большой импульс:

$$q = [\, 2m\,(\Delta E - E_1 - \hbar\omega_0)\,]^{1/2} \tag{8.2}$$

в плоскости квантовой ямы. Значительно бóльшая величина взаимодействия электронов с оптическими фононами, нежели с акустическими, определяет малость τ_q по сравнению с временем захвата из центра.

Во-вторых, τ_q немонотонным, осциллирующим образом зависит от параметров ямы. Это связано с описанными в разделе 4.3 свойствами волновой функции электронов в делокализованных состояниях над квантовой ямой ψ_E. Если яма не является резонансной, то, согласно формуле (4.7), амплитуда этой волновой функции в непосредственной окрестности ямы при малой энергии электрона весьма мала. Матричный элемент электрон-фононного взаимодействия между ψ_1 и ψ_E, так же как и рассматривавшийся в разделе 4.3 электрон-фотонный матричный элемент, будет весьма мал. Собственно, τ_q будет относительно велико. Для резонансных квантовых ям (4.8) вероятность захвата возрастает, т. е. τ_q падает.

Фотопроводимость рассматриваемой структуры, так же как и обычного фоторезистора, определяется произведением трех факторов: скорости оптической генерации, которая в свою очередь пропорциональна коэффициенту поглощения α, времени жизни в делокализованном состоянии τ_q и эффективной подвижности в нем $\mu_{эф}$, которая, очевидно, должна быть пропорциональна квантово-механическому коэффициенту прохождения электрона над квантовой ямой. Первый и третий факторы максимальны для резонансных квантовых ям, а τ_q, напротив, минимально для них. Однако анализ показывает, что совокупное действие всех факторов оказывается таковым, что фотоприемники на квантовых ямах будут иметь лучшие параметры в случае резонансных ям.

Для самой распространенной гетеросистемы $GaAs–Al_xGa_{1-x}$ с $x = 0.2$—0.25 условие резонанса выполняется для ям с толщиной, кратной 40—45 Å. Если $a = 40$—45 Å, то диапазон фоточувствительности структуры лежит в области длин волн порядка 8 мкм, соответствующей одному из окон атмосферной прозрачности и потому очень важной для практических применений. Приемники на основе квантовых ям могут составить конкуренцию фоточувствительным структурам на основе твердых растворов CdHgTe — важнейшему типу приемников для данного спектрального диапазона. Основным достоинством структур на квантовых ямах является бóльшая стабильность и меньший разброс параметров, что особенно важно для матричных фоточувствительных структур.

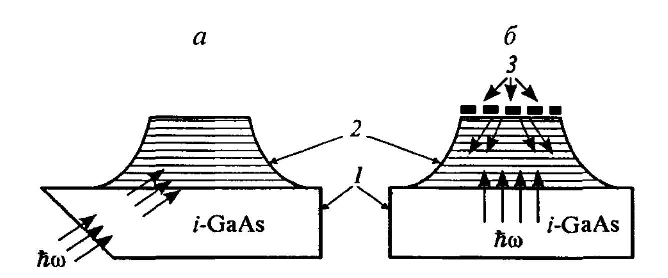


Рис. 8.4. Способы ввода излучения в фотоприемник с квантовыми ямами.

а — через скошенный торец подложки; *б* — с помощью дифракционной решетки. *1* — подложка; *2* — фоточувствительная структура с квантовыми ямами; *3* — дифракционная решетка.

Путем сравнительно небольших изменений состава широкозонных слоев и толщины ямы можно менять положение максимума и ширину полосы фоточувствительности. Последнее обстоятельство связано с тем, что по мере нарушения точного условия резонанса (4.8) спектр фотоионизации квантовой ямы становится более плавным и имеет менее резкий максимум (см. рис 4.3).

В связи с тем что оптическая ионизация квантовых ям может вызываться лишь светом, поляризованным по нормали к квантовым слоям, описанные фотоприемники должны содержать специальные приспособления, поляризующие падающий свет требуемым образом. Есть два основных способа сделать это. Свет может направляться в фоточувствительную структуру под углом через скошенный торец подложки (рис. 8.4, *а*). В другом варианте свет проходит через подложку по нормали, а должную поляризацию приобретает после дифракции на решетке, специально нанесенной на верхнюю поверхность структуры (рис. 8.4, *б*).

Возможно альтернативное решение проблемы поляризации, позволяющее избежать описанных выше конструкционных усложнений. Речь идет о выращивании квантовых структур из полупроводников с анизотропным энергетическим спектром. При наличии анизотропии электрическое поле нормально падающей световой волны, лежащее в плоскости слоев, придает электронам импульс под некоторым углом к этой плоскости. С позиций квантовой механики это означает возможность переходов между различными квантово-размерными уровнями или между уровнем и континуумом состояний над квантовой ямой, что и требуется для работы приемника. На практике для реализации этой идеи чаще всего используют гетероструктуры на основе той же, наиболее освоенной технологически, системы $GaAs–Al_xGa_{1-x}As$, но имеющие не *n*-, а *p*-тип легирования. При этом сложный характер энергетического спектра валентной зоны обеспечивает фоточувствительность при нормальном падении света.

8.5. Лавинные фотодиоды

Лавинные фотодиоды представляют собой фоточувствительные приборы с внутренним усилением, позволяющие получить высокую чувствительность. Основным их недостатком является то, что с лавинным умножением связан дополнительный шум, ограничивающий возможность детектирования слабых сигналов. Уже давно установлено, что для получения низкого уровня шума при большом внутреннем усилении необходимо, чтобы коэффициенты ударной ионизации электронов α и дырок β резко различались между собой.

К сожалению, в большинстве соединений $A^{III}B^{V}$ $\alpha/\beta \cong 1$, что приводит к возрастанию шума при умножении. Поэтому большое практическое значение имеют методы, позволяющие в указанных материалах увеличить α/β. Одним из спрособов сделать это является создание структур типа сверхрешеток, использующих явление ударной ионизации на разрыве энергетических зон [11]. Рассмотрим зонную диаграмму сверхрешеточной структуры в сильном электрическом поле обратно-смещенного *p-i-n*-диода (рис. 8.5).

Пусть мы имеем горячий электрон, ускоряющийся в барьерном слое широкозонного полупроводника. Влетая в узкозонный слой, он резко увеличивает энергию на величину разрыва зоны проводимости ΔE_c. Это эквивалентно тому, что он «видит» энергию ионизации уменьшенной на ΔE_c по сравнению с пороговой энергией в массивном узкозонном полупроводнике. Поскольку коэффициент ударной ионизации α с уменьшением пороговой энергии экспоненциально растет [33], следует ожидать резкого увеличения эффективного значения α. В следующем барьерном слое пороговая энергия увеличивается на ΔE_c, уменьшая тем самым α в этом

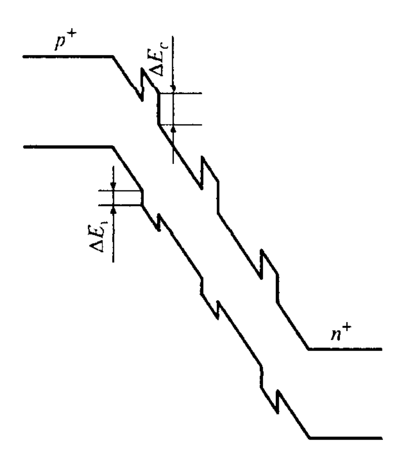


Рис. 8.5. Схема лавинного фотодиода с системой квантовых ям.

слое. Но поскольку $\alpha_1 \ll \alpha_2$ (индексы 1 и 2 относятся соответственно к широкозонному и узкозонному материалам), то экспоненциальный рост α_2 приводит к тому, что и среднее значение $\overline{\alpha}$

$$\overline{\alpha} = \frac{\alpha_1 b + \alpha_2}{a + b} \tag{8.3}$$

значительно увеличивается.

Если (как это, в частности, имеет место в системе GaAs–AlGaAs) разрывы в валентной зоне ΔE_v значительно меньше разрывов в зоне проводимости, то подобный эффект для дырочного коэффициента β будет значительно меньше. Окончательным результатом будет сильное увеличение отношения α/β, что и являлось нашей целью.

8.6. Транзисторы с высокой подвижностью носителей

Одним из самых распространенных видов полупроводниковых приборов являются полевые транзисторы. К их числу принадлежит, в частности, и рассмотренная в разделе 1.3 МДП-структура. Если к инверсионному слою создать два контакта, называемые истоком и стоком и расположенные на расстоянии l друг от друга, то при напряжении между ними, равном V, в канале будет течь ток:

$$I = e\mu n_S Vb/l, \tag{8.4}$$

где b — ширина канала, а μ — подвижность носителей в нем. Изменяя с помощью затвора концентрацию в канале n_S, мы можем осуществлять управление током исток—сток аналогично тому, как в обычном (биполярном) транзисторе напряжение базы управляет током коллектор—эмиттер. Поскольку величина n_S описывается формулой (1.9), то важнейшая характеристика транзистора — крутизна G — в нашем случае определяется выражением

$$G = \frac{dI}{dV_3} = \frac{\varepsilon_d \mu b V}{4\pi d l}. \tag{8.5}$$

Видно, что крутизна полевого транзистора пропорциональна подвижности носителей. Поэтому сразу возникает желание использовать в полевых транзисторах структуры с модулированным легированием, описанные в разделе 1.3 и обладающие гигантской подвижностью носителей. Для этого требуется решить задачу об управлении концентрацией двумерных носителей в этих структурах. Она решается путем создания контакта Шоттки на широкозонном легированном слое. Если толщина этого слоя достаточно мала, то области обеднения вблизи гетероперехода и контакта Шоттки пере-

крываются и все доноры широкозонного слоя оказываются ионизованными. В результате он играет роль диэлектрика, гетероструктура становится аналогом МДП-структуры и концентрация двумерных электронов линейно зависит от напряжения на затворе (контакте Шоттки).

Созданные таким образом полевые транзисторы часто обозначают английский аббревиатурой HEMT (High Electron Mobility Transistor — транзистор с высокой подвижностью электронов) [13]. На самом деле полностью использовать преимущества высокой подвижности двумерного газа и получить транзисторы с крутизной, во много раз большей, чем у обычных МДП-транзисторов, не удается. Причина в том, что в реальных приборах для получения высокого быстродействия и высокой плотности интеграции в БИС расстояние между истоком и стоком l делается весьма малым (~1 мкм). При этом напряженность поля в канале настолько велика, что дрейфовая скорость электронов уже не пропорциональна полю (с коэффициентом, равным подвижности), а выходит на насыщение. В результате крутизна G не описывается формулой (8.5), а слабее зависит от подвижности носителей.

Тем не менее использование структур с модулированным легированием приводит к некоторому повышению крутизны полевых транзисторов. При температуре 77 К она может быть доведена до значений порядка 500 мСм/мм.

Два других важных параметра — характерное время переключения и энергетические затраты на одно переключение транзистора — также могут быть уменьшены по сравнению со стандартными полевыми транзисторами на однородном GaAs. В настоящее время ряд фирм уже ведет конкретные разработки быстродействующих компьютеров, использующих в качестве элементной базы полевые транзисторы с модулированным легированием.

8.7. Транзисторы на горячих электронах

К настоящему времени разработаны два типа транзисторов, для работы которых существенное значение имеет тот факт, что электроны, пролетающие через канал или базу, являются горячими, т. е. имеют кинетическую энергию значительно выше равновесной.

В первом типе транзисторов на основе горячих электронов фактически используется структура полевого транзистора с высокой подвижностью носителей заряда (HEMT-транзистора). В таком транзисторе ток течет в канале, образованном двумерным электронным газом. При увеличении электрического поля в канале температура электронов возрастает и может оказаться настолько высокой, что электроны могут с помощью термоэлектронной эмиссии перейти в слой широкозонного твердого раствора, где их

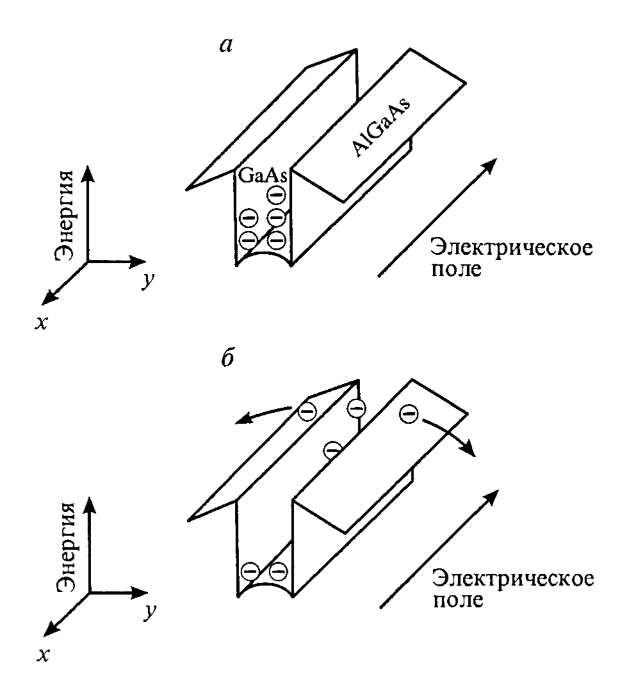


Рис. 8.6. Иллюстрация переноса электронов через канал протекания тока с двумерным электронным газом.

а — в слабом электрическом поле электроны локализованы в одномерной потенциальной яме; *б* — в сильном электрическом поле значительная часть электронов приобретает от поля энергию, достаточную для выхода из потенциальной ямы.

скорость становится малой. Этот эффект иллюстрирует рис. 8.6, на котором представлены энергетические диаграммы канала НЕМТ-транзистора в слабом и сильном электрическом поле. Такой механизм протекания тока может привести к образованию отрицательной дифференциальной проводимости канала. Работа полевого транзистора с отрицательным сопротивлением — ПТОС (Negative Resistance Field-Effect Transistor — NERFET) [59] — основана на этом принципе. Структура такого транзистора представлена на рис. 8.7. Возрастание температуры электронов с ростом напряжения сток—исток приводит к увеличению тока $I_{гор}$, протекающего от канала (эмиттера) через барьер к коллектору, и, следовательно, к уменьшению тока сток—исток, т. е. к отрицательному дифференциальному сопротивлению канала. Основное преимущество такого прибора — возможность реализации более быстродействующих режимов работы, поскольку управление током эмиттер—коллектор в этом случае связано с разогревом электронов. Изменение температуры электронов ограничено наибольшим из двух характерных времен — временем релаксации энергии (порядка 1 пс в GaAs) и временем изменения электрического поля.

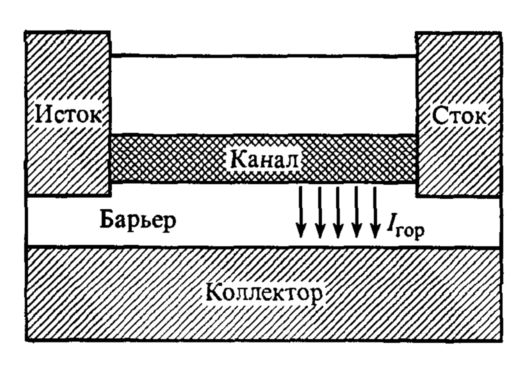


Рис. 8.7. Схематическое изображение структуры ПТОС-транзистора.

Последнее определяется временем пролета электронов через область сильного поля вблизи стока и может быть в несколько раз короче времени их пролета через весь канал, которое ограничивает собственное быстродействие обычных полевых транзисторов. Помимо высокого быстродействия транзисторов на горячих электронах такого типа привлекает возможность построения на их основе новых приборов — с расширенными функциональными возможностями. Например, в четырехэлектродной структуре, которая может быть получена добавлением к ПТОС всего лишь еще одного электрода, подобного стоку или истоку, может быть реализована логическая функция, для выполнения которой на обычных элементах требуется использовать несколько транзисторов.

Транзисторы на горячих электронах второго типа аналогичны биполярным транзисторам, т. е. ток в них протекает не вдоль тонкого слоя (канала), как в полевых транзисторах, а поперек тонкого слоя — через базу. Однако в отличие от биполярных транзисторов они являются униполярными.

Для практики требуются все более быстродействующие приборы, а относительно тяжелая масса дырок является ограничивающим фактором для биполярных транзисторов. Кроме того, электрические напряжения, необходимые для переключения таких приборов, довольно большие — они сравнимы с шириной запрещенной зоны полупроводника. У униполярных транзисторов нет таких проблем — перенос тока в них осуществляется только электронами, электроды разделены потенциальными барьерами, созданными изгибом дна зоны проводимости. Первые униполярные транзисторы — транзисторы с металлической базой — известны с 60-х гг. В таком транзисторе тонкая металлическая база отделена от полупроводниковых эмиттера и коллектора барьерами Шоттки (рис. 8.8, *а*). Повышение быстродействия транзистора определяется возможностью значительного уменьшения толщины базы при сохранении достаточно низкого поперечного сопротивления этого слоя. К сожалению, эту возможность не удалось реализовать. Вы-

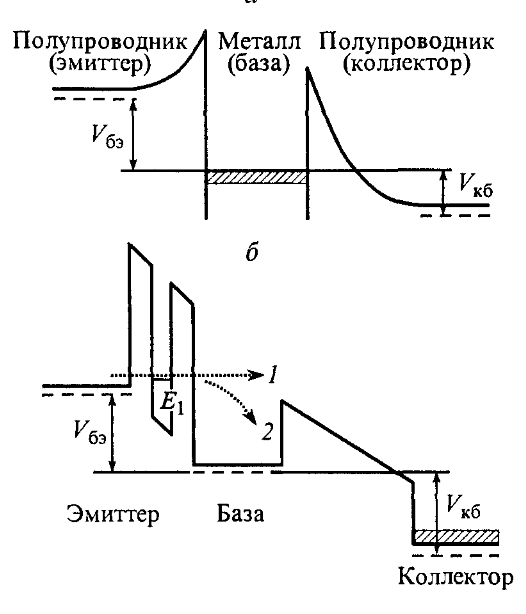


Рис. 8.8. Энергетические диаграммы транзистора с металлической базой (*а*) и транзистора на горячих электронах с резонансным туннелированием (*б*).

1 — горячие электроны; *2* — термализация электронов.

ращивание высококачественных тонких слоев металла на полупроводниковой подложке и тем более слоя полупроводника на металле оказалось практически невыполнимой задачей. В металлическом слое не удается добиться уменьшения рассеяния эмиттированных электронов, чтобы обеспечить достаточную эффективность их переноса через базу в коллектор. Кроме того, два включенных навстречу друг другу барьера Шоттки тоже снижают эту эффективность — из-за квантового отражения электронов барьером коллектор—база. Эти трудности не привели к широкому применению транзисторов с металлической базой, хотя они и выпускались серийно, в том числе и в нашей стране.

В течение 80-х гг. были выполнены многочисленные исследования, направленные на создание подобных транзисторов на основе только полупроводников — без металлической базы. Однако высокое быстродействие с достаточно большим усилением так и не было достигнуто. Высокое легирование базы, необходимое для ее низкого поперечного сопротивления, приводило к сильному рассеянию инжектированных из эмиттера электронов. Поэтому, чтобы достигнуть достаточно высокой эффективности переноса тока через базу, необходимо снижать легирование базы, но это приводит к

снижению быстродействия прибора, поскольку высокое поперечное сопротивление базы увеличивает ее *RC*-характеристику.

Изготовление конкурентоспособного в перспективе транзистора оказалось возможным только после решения проблем, связанных с выбором полупроводникового материала для слоя базы, улучшением технологии нанесения слоев, оптимизацией энергетической структуры транзистора, учетом и использованием квантово-размерных эффектов. Поиски оптимального построения униполярного прибора привели к созданию транзистора на горячих электронах с резонансным туннелированием (Resonant Tunneling Hot Electron Transistor — RHET) [52]. На рис. 8.8, *б* представлена энергетическая структура такого транзистора.

В качестве материала базы в нем используется твердый раствор *n*-InGaAs. В чистом арсениде галлия горячие электроны, как показали исследования, теряют свою энергию на величину ~60 мэВ на расстоянии 200 нм. Ясно, что выгодно увеличивать энергию электронов, чтобы бóльшая их часть достигала коллектора. Однако в GaAs не имеет смысла инжектировать электроны с энергией, большей чем 0.3 эВ, так как межзонное рассеяние в долины с тяжелыми электронами снизит эффективность прибора. Для InGaAs экстремумы с низкой подвижностью находятся значительно выше по энергии. В этот материал могут быть инжектированы электроны с энергией > 0.4 эВ. В этом случае могут быть скомпенсированы потери эффективности за счет термализации электронов — бóльшая часть электронов будет достигать коллектора. В этом приборе в качестве барьера эмиттер—база используется двойная барьерная структура AlGaAs–GaAs или AlGaAs–InGaAs. Электроны инжектируются через такую структуру в базу с помощью резонансного туннелирования — с относительно узким распределением по энергии, что существенно снижает потери, связанные с термализацией электронов. RHET-транзисторы работают пока только лишь при температуре жидкого азота. Однако работы, направленные на улучшение характеристик таких приборов, на повышение их рабочих температур, продолжаются. Это связано с тем обстоятельством, что в транзисторах такого типа есть принципиальная возможность достижения терагерцового диапазона рабочих частот. Горячие электроны двигаются через базу баллистически, т. е. их скорость достигает почти предельных величин. Время пролета через базу может составлять всего лишь несколько десятков фемтосекунд. Быстродействие транзистора ~1 ТГц возможно, даже при учете ограничивающего действия времени пролета электрона через коллектор. Кроме того, баллистическое движение электронов и отсутствие рекомбинации электронов с дырками значительно снижают шумовые характеристики — в этом также большое преимущество таких транзисторов на горячих электронах.

8.8. Резонансно-туннельный транзистор на квантовой точке

Двухбарьерная резонансно-туннельная структура, рассмотренная в главе 5, представляет собой диодную, двухэлектродную структуру. Ее вольт-амперная характеристика (см. рис. 5.8), имеющая участок отрицательного дифференциального сопротивления, определяется только технологией изготовления структуры. Такой резонансно-туннельный диод может быть использован как отдельный прибор при построении электронных схем, так и в качестве элемента более сложных транзисторных структур. В транзисторе на горячих электронах, который мы рассмотрели в предыдущем параграфе, он используется в качестве барьера эмиттер— база. Однако возможно применение резонансного туннелирования и в трехэлектродной структуре. Такой резонансно-туннельный транзистор можно создать, если использовать резонансное туннелирование не через планарную двухбарьерную структуру, а через квантовую точку [60].

Квантовая точка имеет дискретный энергетический спектр. На ее основе можно изготовить резонансно-туннельный диод, если связать ее через туннельно-прозрачные барьеры с двумя электродами. Вольт-амперная характеристика такой структуры будет иметь участок отрицательного дифференциального сопротивления. Однако в отличие от планарной структуры в этом случае оказывается возможным осуществить управление проводимостью структуры. Для этого необходимо иметь способ изменения размеров квантовой точки. В таком случае будет меняться положение энергетических уровней в квантовой точке — появляется принципиальная возможность «включать» и «выключать» механизм резонансного туннелирования. На рис. 8.9 представлено схематическое изображение резонансно-туннельного транзистора на основе квантовой точки, в котором реализуется это условие. Центральный верхний электрод транзистора круглой формы соединяется с нижним электродом через планарную двухбарьерную резонансно-туннельную структуру с двумерным электронным газом в центре. Квантовая точка в этой структуре образуется с помощью третьего электрода — затвора, кольцом окружающего центральный верхний электрод. При подаче на него отрицательного потенциала электроны из области двумерного газа под затвором вытесняются к центру структуры. Таким способом под центральным электродом может быть сформирована квантовая точка, поперечные размеры которой, а следовательно, и положение энергетических уровней в ней определяются величиной отрицательного напряжения на затворе. Сдвиг уровней приводит к изменению условий резонансного туннелирования. Положение участков отрицательного дифференциального сопротивления в вольт-амперной характеристике между цент-

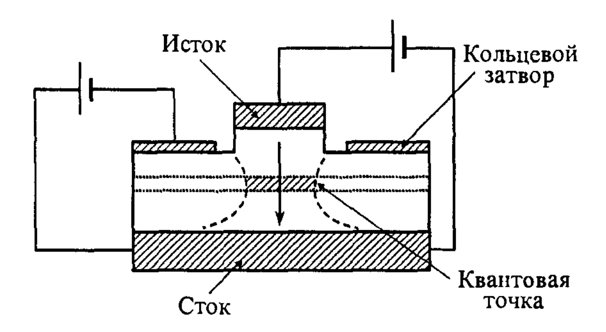


Рис. 8.9. Схематическое изображение структуры резонансно-туннельиого транзистора на основе квантовой точки.

ральным и нижним электродами зависит от напряжения на затворе — такой прибор имеет более широкие функциональные возможности, чем просто резонансно-туннельный диод.

На работе резонансно-туннельного транзистора, к сожалению, отрицательно сказывается наличие неконтролируемых примесей и дефектов в области квантовой точки и туннельных барьеров. Различное положение примесных атомов в области квантовой точки для разных транзисторов приводит из-за искажений локального потенциала к значительному разбросу характеристик транзисторов. Кроме того, через электронные атомные уровни примеси тоже может происходить резонансное туннелирование — вольт-амперная характеристика транзистора будет иметь в этом случае пики, положение которых не зависит от напряжения на затворе. Однако резонансно-туннельные транзисторы потребляют очень малую мощность на одно переключение — в этом их преимущество.

8.9. Приборы на основе баллистического транспорта

Эффект преломления направления движения баллистически движущихся электронов на границе раздела двух областей двумерного электронного газа с различной концентрацией носителей заряда, описанный в главе 5, можно использовать для создания принципиально нового типа приборов. С помощью металлических затворов специальной формы можно создавать такие потенциальные профили в плоскости двумерного электронного газа, которые будут играть роль «линз» или «призм» для баллистических электронов. Открывается возможность конструирования электронных приборов, подобных оптическим устройствам.

На рис. 8.10 схематически представлен пример такой конструкции — коммутатор пучка электронов, инжектируемых в его рабо-

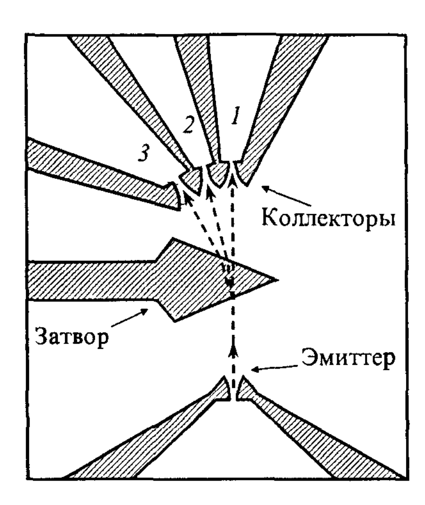


Рис. 8.10. Схематическое изображение коммутатора пучка баллистических электронов.

чую область эмиттером, между тремя коллекторами (*1*—*3*) [60]. Размеры активной области прибора меньше длины свободного пробега электронов в двумерном электронном газе, расположенном в плоскости рисунка, поэтому они пролетают расстояние между эмиттером и коллекторами без столкновений — баллистически. Коммутация осуществляется с помощью затвора треугольной формы, помещенного между эмиттером и коллекторами. При подаче на затвор положительного напряжения концентрация электронов под ним из-за эффекта Шоттки будет больше, чем в свободном электроном газе, — эта область будет играть роль «оптической» призмы для пучка инжектированных электронов.

Для работы такого прибора необходимо иметь пучок инжектированных электронов, двигающихся параллельно. Эту задачу выполняет эмиттер — квантовый точечный контакт специальной формы (рис. 8.11). На оба затвора контакта подается отрицательный потенциал, под ними образуется область обеднения. Узкое пространство между затворами образует квантовый точечный контакт, имеющий форму рупора. Такой контакт способен коллимировать (сводить в параллельный поток) пучок электронов. Это явление известно в микроволновой технике, где плавно сужающийся рупор может быть использован для коллимации микроволнового луча. Как видно из рис. 8.11, электрон, попадающий в рупор под достаточно большим углом к его оси, покидает его с траекторией, близкой к оси рупора. Коллимация осуществляется за счет многократных отражений от стенок рупора. Коллекторы рассматриваемого нами прибора тоже представляют собой квантовые точечные контакты такой же формы — своего рода приемные рупорные «антенны».

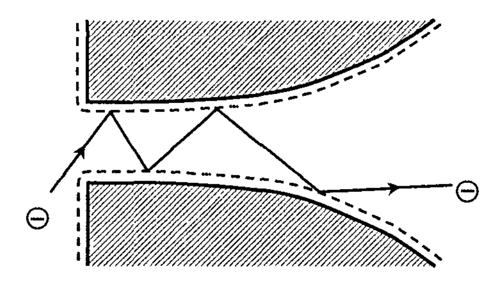


Рис. 8.11. Коллимация пучка электронов с помощью квантового точечного контакта специальной формы.

При нулевом напряжении на затворе электронной призмы не существует и пучок электронов, не отклоняясь, попадает в первый коллектор. Как только к затвору будет приложен положительный потенциал, проявится эффект преломления направления движения электронов при пересечении ими области двумерного электронного газа с большей концентрацией носителей под затвором. При некотором напряжении на затворе ток будет максимальным во втором коллекторе, а еще при большем — в третьем коллекторе. Таким образом, подавая соответствующее напряжение на затвор, мы можем направлять пучок электронов в любой из трех коллекторов. Такой прибор заменяет собой несколько обычных транзисторов, которые пришлось бы использовать для построения подобного коммутатора. Более того, поскольку электроны пересекают рабочую область коммутатора без столкновений, можно с помощью одного затвора коммутировать одновременно два или более электронных пучка. Для этого необходимо только добавить соответствующее число эмиттеров и коллекторов. Электронные пучки не будут взаимодействовать друг с другом. Такой прибор является ярким примером построения на основе квантово-размерных структур новых приборов с расширенными функциональными возможностями.

8.10. Одноэлектронный транзистор

На основе явления кулоновской блокады, речь о котором шла в разделе 7.3, возможно создание целого класса новых приборов — одноэлектронных транзисторов и устройств на их основе [60—65]. На рис. 8.12, *а* представлен один из вариантов одноэлектронного транзистора в структуре с двумерным электронным газом. При отрицательном напряжении на затворах G_1, G_2 и G_4 вблизи них образуется область, обедненная носителями заряда, в результате в центре структуры создается проводящий островок очень малых размеров (квантовая точка), с помощью туннельного эффекта связанный с резервуарами истока и стока. Связь квантовой точки с основным за-

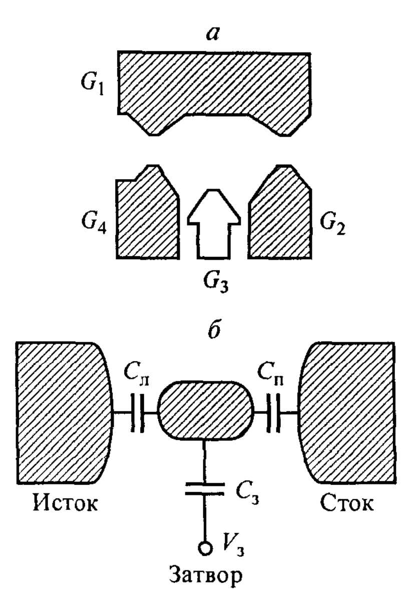


Рис. 8.12. Одноэлектронный транзистор в структуре с двумерным электронным газом (*а*) и его схематическое изображение (*б*).

твором транзисторной структуры G_3 является чисто емкостной — расстояние между ними достаточно велико, чтобы исключить вероятность туннельных переходов. На рис. 8.12, *б* схематически изображена структура одноэлектронного транзистора. Затвор используют для изменения электрохимического потенциала электронов в центральном электроде с помощью наведенного квазизаряда $\Delta Q = C_з V_з$. В отличие от заряда, связанного с туннелированием электронов от истока к стоку через квантовую точку, этот заряд может изменяться непрерывно, так как это поляризационный заряд.

С помощью концепции квазизаряда мы можем определить условия переключения одноэлектронного транзистора. Энергия центрального электрода по-прежнему определяется выражением $E = q^2/2C$, где q — его полный заряд, а C — электрическая емкость островка: $C = C_л + C_п + C_з$. Если на центральном электроде находится N дополнительных электронов, то его заряд равен $q = -Ne + \Delta Q = -Ne + C_з V_з$, так что полная энергия задается выражением

$$E = \frac{(-Ne + C_з C_з)^2}{2C}. \tag{8.6}$$

Если между истоком и стоком приложено небольшое напряжение смещения, так что будет иметь место режим кулоновской блокады, то ток через структуру течь не может, так как энергия центрального электрода увеличивается с каждым дополнительным электроном:

$E(N+1) > E(N)$. Однако, подстраивая напряжение на затворе, мы можем добиться выполнения условия $E(N+1, V_з + \Delta V_з) = E(N, V_з)$. В этом случае электрон может туннелировать через структуру и ток через транзистор будет течь. Другими словами, из этого условия видно, что электроны могут туннелировать в центральный электрод всякий раз, когда напряжение на затворе изменяется на величину $\Delta V_з = e/C_з$, т. е. проводимость одноэлектронного транзистора будет осциллировать с увеличением напряжения на затворе с периодом $\Delta V_з$ (рис. 8.13).

В одноэлектронном транзисторе с одним центральным электродом может оказаться существенным явление, которое способно нарушить нормальную работу транзистора. Это явление называют сотуннелированием (co-tunneling) или макроскопическим квантовым туннелированием. Дело в том, что помимо последовательного туннелирования электрона сначала из истока в квантовую точку, а затем из нее в сток существует малая, но конечная вероятность непосредственного переноса заряда из истока в сток. Этот перенос может осуществляться двумя механизмами. В первом из них электрон может пересечь область центрального электрода с помощью туннельного эффекта через виртуальное состояние или посредством резонансного туннелирования. При втором механизме один электрон туннелирует в центральный электрод из истока, а второй тут же покидает эту область, переходя на сток. При этом в центральном электроде образуется электронно-дырочная пара. Сотуннелирование, хотя и является эффектом следующего порядка малости в сравнении с единичным актом туннелирования, приводит к конечной проводимости транзистора даже в условиях кулоновской бло-

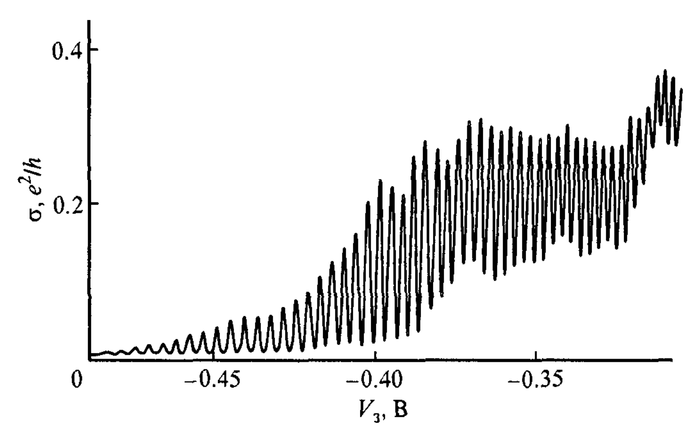


Рис. 8.13. Проводимость одноэлектронного транзистора, представленного на рис. 8.12, *а*. Каждый пик проводимости соответствует удалению (с ростом отрицательного смещения) очередного электрона из квантовой точки, причем при смещении −0.4 В число электронов в точке порядка 100 [53].

кады. Для подавления сотуннелирования используют структуры с двумя и более центральными электродами, отделенными друг от друга туннельно-прозрачными барьерами (MTJ — Multi-Tunnel Junctions). В таких MTJ-транзисторах вероятность мгновенного переноса заряда из истока в сток практически равна нулю.

В настоящее время одноэлектронные транзисторы работают только при низких температурах, хотя в ряде структур одноэлектронные эффекты наблюдались и при комнатной температуре. Для создания приборов, надежно работающих при комнатной температуре, необходимо решать проблемы, связанные с уменьшением размеров квантовых точек и уменьшением емкости затвора. Работы, направленные на оптимизацию структуры одноэлектронных транзисторов, интенсивно проводятся во многих лабораториях мира. Такие транзисторы открывают широкие перспективы одноэлектронной цифровой логики, в которой бит информации может быть представлен только лишь одним электроном. Кроме того, одноэлектронные транзисторы имеют очень высокое быстродействие — по оценкам, до нескольких сотен терагерц при чрезвычайно малой потребляемой мощности. Одноэлектроника является одной из самых перспективных и быстро развивающихся областей наноэлектроники.

8.11. Устройства на основе одноэлектронных транзисторов

Стандарт силы тока (электронный «насос»). В качестве стандартов единицы сопротивления (ом) и единицы напряжения (вольт) могут служить соответственно квантовый эффект Холла и эффект Джозефсона. Стандарт единицы силы электрического тока (ампер) был основан на экспериментах, дающих значительно меньшую точность. С помощью одноэлектронных эффектов удается определить единицу силы тока со значительно большей точностью.

Структура из квантовой точки, туннельными переходами связанная с истоком и стоком, может служить стандартом силы тока [66, 67]. Планарную квантовую точку выделяют из области двумерного электронного газа с помощью системы затворов (рис. 8.14). Затвор *3* аналогичен затвору одноэлектронного транзистора и служит для изменения электростатической энергии квантовой точки. Затворы *2* и *4* совместно с затвором *1* определяют высоту, а значит, и прозрачность туннельных барьеров, через которые происходит перенос электронов. Принципиальное отличие такой структуры от обычного одноэлектронного транзистора состоит в том, что с помощью напряжения на затворах *2* и *4* высоту левого и правого туннельных барьеров можно изменять независимо. Между истоком и

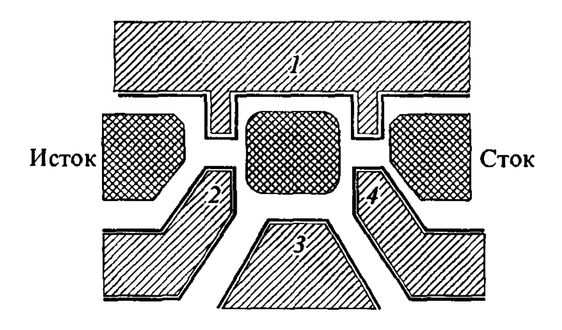


Рис. 8.14. Расположение поверхностных затворов (*1*—*4*) на структуре с двумерным электронным газом для создания стандарта силы тока.

стоком структуры прикладывают небольшое напряжение смещения V_c — такое, чтобы в квантовую точку можно было поместить только один электрон: $eV_c \geqslant E(N+1) - E(N)$ (рис. 8.15, *а*). На затворы 2 и *4* подают в противофазе переменное напряжение с частотой *f*. Таким образом, когда повышается один барьер, другой в этот момент понижается. Прохождение электрона от левого электрода к правому через квантовую точку в течение периода изменения напряжения на затворах 2 и 4 продемонстрировано на рис. 8.15, *а*—*г*. Через пониженный левый барьер электрон проникает в квантовую точку, но выйти из нее в правый электрод не может из-за низкой вероятности туннелирования через высокий правый барьер (рис. 8.15, *б*). Только лишь при следующем полупериоде переменного напряжения на затворах электрон покидает квантовую точку и переходит в правый электрод (рис. 8.15, *г*). Таким образом, за один период изменения напряжения на затворах 2 и *4* через квантовую точку проходит только один электрон. По

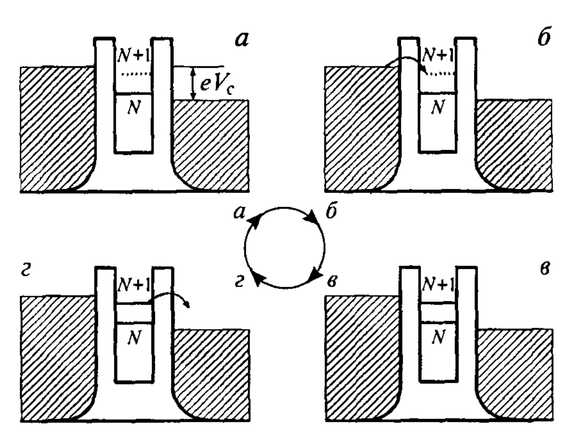


Рис. 8.15. Схема прохождения одного электрона через квантовую точку при периодической модуляции высоты туннельных барьеров.

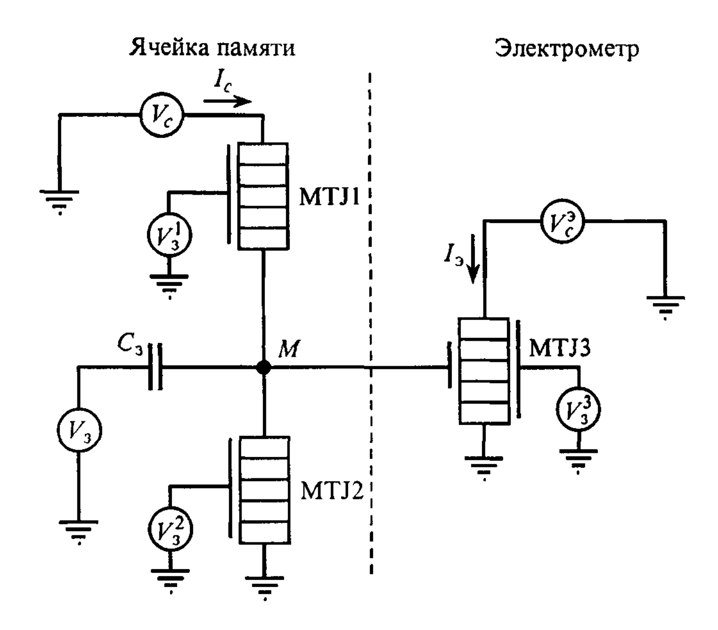


Рис. 8.16. Ячейка памяти на основе одноэлектронных транзисторов.

этой причине такую структуру называют одноэлектронным «насосом». Ток, протекающий через точку, будет равен $I = ef$. Величина заряда электрона известна с очень высокой степенью точности, измерение частоты сводится к измерению интервалов времени, что можно сделать тоже достаточно точно. В результате оказывается возможным измерить силу тока с точностью, значительно превосходящей точность других стандартов тока.

Ячейка памяти на основе одноэлектронных транзисторов. Одноэлектронный транзистор может находиться в двух состояниях — либо в состоянии кулоновской блокады, либо в проводящем состоянии. Этот факт позволяет создавать на его основе устройства, которые могут служить ячейкой компьютерной памяти. Очень малые размеры одноэлектронных транзисторов при высоком быстродействии и низкой потребляемой мощности делают такое их применение очень перспективным. В качестве примера рассмотрим одну из возможных конструкций ячейки памяти на основе одноэлектронных эффектов (рис. 8.16) [68]. В качестве активных элементов в ней используются три одноэлектронных транзистора на основе множественных туннельных переходов (MTJ), о которых мы уже упоминали выше. Основными действующими элементами ячейки являются транзисторы MTJ1 и MTJ3. Транзистор MTJ2 играет вспомогательную роль и при работе ячейки находится в выключенном состоянии (в режиме кулоновской блокады).

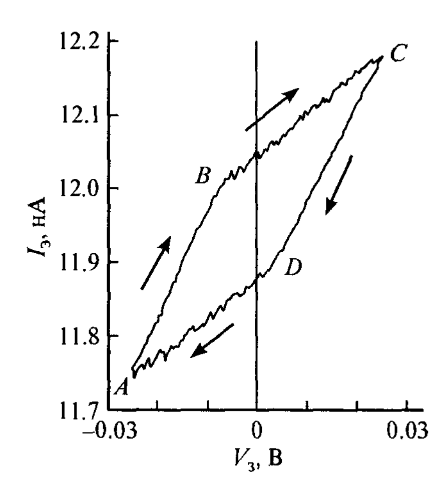


Рис. 8.17. Гистерезис в зависимости тока электрометра при различных направлениях изменения напряжения на затворе ячейки памяти.

Транзистор MTJ3 имеет два затвора. С помощью напряжения $V_з^3$ на его правом затворе рабочую точку транзистора устанавливают на пороге его открывания, поэтому очень малые изменения потенциала второго затвора, соединенного с узлом запоминания *M*, приводят к увеличению или уменьшению тока $I_э$ через MTJ3. Таким образом, этот транзистор выполняет функции электрометра — с его помощью можно регистрировать малые изменения заряда в точке *M*.

На рис. 8.17 представлена зависимость тока электрометра $I_э$ от напряжения на основном затворе ячейки $V_з$. Гистерезис этой зависимости (направления изменения $V_з$ указаны стрелками) связан с различным состоянием транзистора MTJ1 на разных участках его петли. На участке *AB* он закрыт, находится в режиме кулоновской блокады. Скорость накопления заряда в точке *M* большая, она определяется параметрами емкостного делителя, образованного емкостью $C_з$ и емкостями транзисторов MTJ1 и MTJ2. Однако при потенциале в узле *M*, соответствующем точке *B* кривой, транзистор MTJ1 открывается — выходит из режима кулоновской блокады. В результате скорость накопления заряда в узле замедляется — заряд стекает через открытый транзистор. При смене направления изменения $V_з$ в точке *C* кривой транзистор MTJ1 закрывается и скорость изменения заряда снова будет определяться только емкостным делителем. В точке *D* кривой транзистор MTJ1 снова открывается — при другой полярности снятия кулоновской блокады. Изменение потенциала в точке *M* снова замедляется. Таким образом, при $V_з = 0$ ячейка имеет два устойчивых состояния, которые могут быть использованы для записи бита информации.

8.12. Квантово-точечные клеточные автоматы и беспроводная электронная логика

Потребности в разработке логических устройств для нанокомпьютеров с очень высокой плотностью логических элементов и с максимально возможно низким потреблением энергии на одно переключение привели к предложениям использовать в логических элементах проводящие островки очень малого размера — квантовые точки [69, 70]. В таких приборах для реализации вычислений логических булевых функций [71, 72] используют массивы связанных взаимодействующих квантовых точек. Эти новые приборы называют квантово-точечными клеточными автоматами (QCA — Quantum Cellular Automata).

Основу прибора составляет ячейка, состоящая из четырех или пяти квантовых точек. На рис. 8.18 представлена ячейка из пяти квантовых точек: четыре точки расположены в углах квадрата, а одна — в его центре. В ячейку при помощи внешнего напряжения через дополнительный электрод вводятся два избыточных электрона, и ячейка приобретает электрический заряд. Квантовые точки в ячейке располагаются таким образом, что возможно туннелирование только через центральную точку. Из-за электростатического отталкивания между избыточными электронами вся система будет иметь минимальную энергию только в том случае, если электроны расположатся как можно дальше друг от друга, т. е. в углах квадрата, соединенных диагональю. Поскольку таких возможных положений всего два, то система имеет всего два устойчивых состояния (две поляризации), и, следовательно, одно из этих состояний можно считать логической единицей («1»), а второе — логическим нулем («0»). При переходе системы из одного устойчивого состояния в другое меняются поляризация системы и распределение электрических полей вокруг ячейки. С помощью дополнительных электродов, связанных с ячейкой емкостной связью, можно навязать ячейке необходимое состояние и перевести ее в состояние «1» или «0». Если рядом с первой ячейкой расположить вторую (в которой также находится два дополнительных электрона), то электростатическое поле первой ячейки заставит электроны располагаться так, чтобы обеспечить минимум электростатической энергии всей системы (рис. 8.18). Составляя комбинации из расположенных разным образом ячеек, можно реализовать разнообразные логические функции и выполнить необходимые логические преобразования и вычисления. На рис. 8.19 представлен пример комбинации ячеек, при которой состояние на выходе определяется большинством состояний на входе (логическая функция «Majority»). Предложены разнообразные комбинации ячеек для реализации логических операций. На основе таких элементов возможно создание нанокомпью-

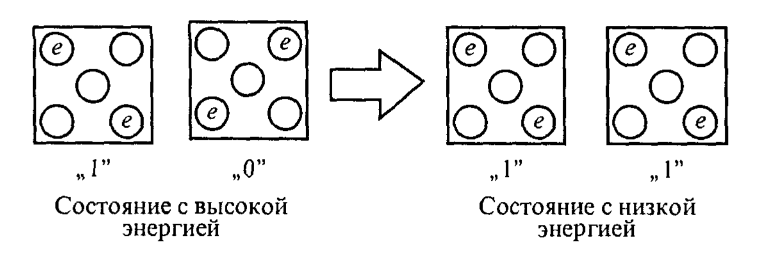


Рис. 8.18. Различные конфигурации ячеек квантово-точечных автоматов.

тера. Важно отметить, что взаимное расположение ячеек обеспечивает передачу логического сигнала без перемещения зарядов вдоль цепочки — в бестоковом режиме, только за счет передачи вдоль цепочки состояния поляризации.

Преимущества логических устройств на основе квантово-точечных клеточных автоматов состоят в том, что по сравнению с аналогичными устройствами на основе полевых транзисторов требуется значительно меньший объем активной области. Например, полный сумматор на основе клеточных автоматов с размером точки 20 нм можно расположить на площади около 1 мкм^2, в то время как такую же площадь занимает всего лишь один полевой транзистор. Для построения такого же сумматора на основе полевых транзисторов требуется примерно 40 транзисторов. Если еще учесть области коммутации транзисторов между собой, а они, как известно, занимают объем, сравнимый или даже превосходящий объем, занятый активными приборами, то преимущества использования клеточных автоматов становятся очевидными.

Принципиальная возможность функционирования логических элементов на основе клеточных автоматов была продемонстрирована при T = 75 мК на примере ячейки, изготовленной при помо-

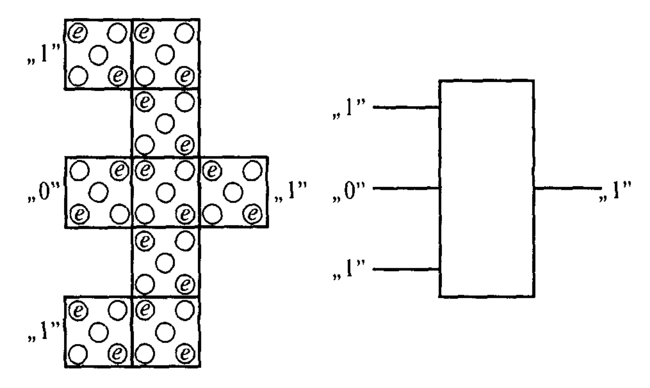


Рис. 8.19. Комбинации ячеек квантово-точечного автомата, при которых состояние на выходе определяется большинством состояний на входе.

щи стандартной электронно-лучевой литографии из алюминиевых островков на поверхности окисленной кремниевой пластины [73]. Площадь прибора составляла величину $\sim 50 \times 50$ нм^2.

Вычислительный процесс в приборах на основе клеточных автоматов осуществляется при переходе всей совокупности ячеек в состояние с минимальной энергией — в основное состояние. Поскольку сложные вычислительные устройства должны содержать большое число ячеек, то состояние с минимальной энергией можно получить разными способами. Это может приводить к ошибкам в вычислениях. Кроме того, такие системы чувствительны к внешним воздействиям и поэтому требуют строгого контроля внешних условий. При повышении температуры вычислительный процесс может быть разрушен. Для ячеек, у которых размер одной квантовой точки ~ 20 нм, изменение энергии при перезарядке точки составляет величину ~ 1 мэВ (примерно $1/20\ kT$ при комнатной температуре). Так же как и для одноэлектронного транзистора, рабочую температуру нужно повышать за счет уменьшения размеров ячейки (и соответственно каждой квантовой точки). Существует еще одна проблема, которая должна быть решена для успешной работы устройств на основе клеточных автоматов. Она состоит в том, что поскольку электростатическое поле ячейки влияет на соседние ячейки как в направлении выхода устройства, так и в направлении входа, то из-за случайных воздействий возможно распространение сигнала не только от входа к выходу, но и наоборот. Для устранения этого недостатка предложены устройства на основе квантовых точек, у которых направление передачи сигнала определяется внешним электрическим полем [74]. Практическое изготовление устройств на основе клеточных автоматов находится в самой начальной стадии и требует разрешения целого ряда проблем, в основном технологических.

ПРИЛОЖЕНИЕ

ВОПРОСЫ И ЗАДАЧИ

1. Получить оценку предельной толщины пленки, при которой возможно наблюдение квантово-размерных явлений, если подвижность электронов в пленке 10^4 см^2/(В · с).

2. Какова предельная толщина пленки, при которой возможно наблюдение квантово-размерных явлений при комнатной температуре, если эффективная масса носителей $m = 0.1m_0$?

3. Электрон с массой m находится в прямоугольной потенциальной яме шириной a с бесконечно высокими стенками. Показать, что собственные значения энергии частицы и ее нормированные собственные функции в области $0 < z < a$ имеют вид $E_N = \dfrac{\hbar^2\pi^2N^2}{2ma^2}$, $\psi_N(z) = \sqrt{\dfrac{2}{a}}\sin\dfrac{\pi Nz}{a}$, $N = 1, 2, 3,...$

4. Для прямоугольной квантовой ямы шириной a и глубиной U получить уравнение для определения значений энергии связанных состояний. Определить число связанных состояний в яме. Найти условие, при котором расстояние по шкале энергий от вершины барьера до нижнего уровня в яме равно заданной величине E_0.

5. Решить предыдущую задачу в случае разных эффективных масс электронов в барьере и яме — $m_б$ и $m_я$.

Указание: воспользоваться граничным условием $\left.\dfrac{1}{m_б}\dfrac{d\psi_б}{dz}\right|_{\substack{z=0\\z=a}} = \left.\dfrac{1}{m_я}\dfrac{d\psi_я}{dz}\right|_{\substack{z=0\\z=a}}$.

6. Найти волновые функции и уровни энергии в треугольной квантовой яме, которая слева в точке $z = 0$ ограничена бесконечно высокой стенкой и справа потенциалом $V = \alpha z$.

Указание: ввести константы $l = \left(\dfrac{\hbar^2}{2m\alpha}\right)^{1/3}$ и $\lambda = l^2\dfrac{2mE}{\hbar^2}$ и сделать замену переменной $z = l(\xi + \lambda)$.

7. Найти уравнение для определения энергии связанных состояний в яме, ограниченной справа в точке $z = 0$ бесконечно высокой стенкой, а справа потенциалом $V(z) = -V_0 e^{-\alpha z}$.

Указание: сделать сначала замену переменной $t = e^{-\frac{\alpha}{2}z}$, а затем $\xi = \sqrt{\frac{8mV_0}{\hbar^2\alpha^2}}\,t$.

8. Поверхность тонкой монокристаллической пленки кремния имеет ориентацию (100). Рассчитайте плотность электронных состояний в таком двумерном электронном газе. Что изменится при ориентации поверхности (111)?

9. Получить правила отбора для межзонных оптических переходов в прямоугольных квантовых ямах.

Указание: см. *Шик А. Я.* Письма в ЖТФ. 1979. Vol. 5. P. 869.

10. Получить выражение для экранированного потенциала точечного заряда в двумерном электронном газе.

Указание: решить уравнение Лапласа в цилиндрической системе координат. Заряд в области двумерного электронного газа использовать в качестве граничного условия.

11. Найти выражение для релаксации заряда в области двумерного электронного газа.

Указание: использовать фурье-преобразование потенциала и концентрации. Рассмотреть затем частный случай профиля начального возмущения концентрации в виде лоренциана: $n(x,0) = \frac{Na}{x^2 + a^2}$.

12. Решить уравнение Шредингера для электронов в двумерном электронном газе, находящемся в плоскости $z = 0$ во внешнем магнитном поле $\mathbf{H} = (0, 0, H_z)$ и электрическом поле $\mathbf{F} = (F_x, 0, 0)$.

13. Найти связь между концентрацией электронов и уровнем Ферми для вырожденного одномерного электронного газа.

Указание: использовать квадратичный закон дисперсии энергии электронного газа.

ЛИТЕРАТУРА

1. *Шик А. Я.* Двумерные электронные системы (учебное пособие). СПб.: Изд-во СПбГТУ, 1993; *Shik A. Ya.* Quantum wells. Physics and electronics of two-dimensional systems. Singapore, New Jersey, London, Hong Kong: World Scientific, 1998.
2. *Демиховский В. Я.* Квантовые ямы, нити, точки. Что это такое? // Соросовский образовательн. журн. 1997. № 5. С. 80.
3. *Тавгер Б. А., Демиховский В. Я.* Квантовые размерные эффекты в полупроводниковых и полуметаллических пленках // УФН. 1968. Т. 96, вып. 1. С. 61.
4. *Андо Т., Фаулер А., Стерн Ф.* Электронные свойства двумерных систем. М.: Мир, 1985.
5. *Шик А. Я.* Полупроводниковые структуры с дельта-слоями // ФТП. 1992. Т. 26. С. 1161.
6. *Шик А. Я.* Квантовые нити // Соросовский образовательн. журн. 1997. № 5. С. 87.
7. *Sawada A., Usagawa T., Ho S., Yamaguchi K.* Possible new structure for one-dimensional electron-gas systems by interface bending of AlGaAs/GaAs heterojunctions // Appl. Phys. Lett. 1992. Vol. 60. P. 1492.
8. *Gaponenko S. V.* Optical properties of semiconductor nanocrystals. Cambridge University Press, 1998.
9. *Ekimov A.* Growth and optical properties of semiconductor nanocrystals in a glass matrix // J. Luminescence. 1996. Vol. 70. P. 1.
10. *Шик А. Я.* Сверхрешетки — периодические полупроводниковые структуры // ФТП. 1974. Т. 8. С. 1841.
11. *Херман М.* Полупроводниковые сверхрешетки. М.: Мир, 1989.
12. *Мусихин С. Ф., Бакуева Л. Г., Ильин В. И. и др.* Оптические и электрические свойства сверхрешеток Фибоначчи PbS—C, полученных методом импульсного лазерного напыления // ФТП. 1995. Т. 29. С. 474.
13. Молекулярно-лучевая эпитаксия и гетероструктуры / Ред. Л. Ченг, К. Плог. М.: Мир, 1989.
14. *Готра З. Ю.* Технология микроэлектронных устройств. М.: Радио и связь, 1991.
15. *Nötzel R.* Self-organized growth of quantum-dot structures // Semicond. Sci. Technol. 1996. Vol. 11. P. 1365.
16. *Леденцов Н. Н., Устинов В. М., Щукин В. А. и др.* Гетероструктуры с квантовыми точками: получение, свойства, лазеры // ФТП. 1998. Т. 32. С. 385.
17. *Shchukin V. A., Bimberg D.* Spontaneous ordering of nanostructures on cristal surfaces // Rev. Mod. Phys. 1999. Vol. 71. P. 1125.

18. *Daruka I., Barabási A.-L.* Dislocation-free islands formation in heteroepitacsial grown: a study at equilibrium // Phys. Rev. Lett. 1997. Vol. 79. P. 3708.
19. *Ландау Л. Д., Лифшиц И. М.* Теория упругости. М.: Наука, 1987.
20. *Марченко В. И., Паршин А. Я.* О свойствах поверхности кристаллов // ЖЭТФ. 1980. Т. 79. С. 257.
21. *Красильник З. Ф., Новиков А. В.* Оптические свойства напряженных гетероструктур на основе $Si_{1-x}Ge_x$ и $Si_{1-x-y}Ge_xC_y$ // УФН. 2000. Т. 170. С. 338.
22. *Хачатурян А. Г.* Теория фазовых превращений и структура твердых растворов. М.: Наука, 1974.
23. *Montemerlo M. S., Love J. C., Opiteck G. J. et al.* Technologies and designs for electronic nanocomputers // http:/www.mitre.org/research/nanotech/list_of_articles.html.
24. *Рыков С. А.* Сканирующая зондовая микроскопия полупроводниковых материалов и наноструктур / Под ред. А. Я. Шика. СПб.: Наука, 2001.
25. *Hansma P. K., Elings V. B., Marti O., Bracker C. E.* Scanning tunneling microscopy and atomic force microscopy // Science. 1988. Vol. 242. P. 209.
26. *Jung T. A., Schlitter R. R., Gimzewski J. K. et al.* Controlled room-temperature positioning of individual molecules: molecular flexure and motion // Science. 1996. Vol. 271. P. 181.
27. *Wendel M., Kühn S., Lorenz H. et al.* Nanolitography with an atomic force microscope for integrated fabrication of quantum electronic devices // Appl. Phys. Lett. 1994. Vol. 65. P. 1775.
28. *Irmer B., Kehrle M., Lorenz H., Kotthaus J. P.* Nanolitography by non-contact AFM-induced oxidation: fabrication of tunneling barries sutable for single-electron devices // Semicond. Sci. Technol. 1998. Vol. 13. P. A79.
29. *Whitesides G. M.* Self-assembling materials // Scientific American. 1995. September. P. 146.
30. *Haug H., Koch S. W.* Quantum theory of the optical and electronic properties of semiconductors. 3th ed. Singapore: World Scientific, 1994.
31. Shi*k A.* Excitons and impurity centers in thin wires and in porous silicon // J. Appl. Phys. 1993. Vol. 74. P. 2951.
32. *Кейдыш Л. В.* Кулоновское взаимодействие в тонких пленках полупроводников и полуметаллов // Письма в ЖЭТФ. 1979. Т. 29, вып. 11. С. 716.
33. *Зеегер К.* Физика полупроводников. М.: Мир, 1977.
34. *Петров А. Г., Шик А. Я.* Поглощение света дырками в квантовых ямах // ФТП. 1994. Т. 28. С. 2185.
35. *Иродов И. Е.* Задачи по квантовой физике. М.: Высшая школа, 1991.
36. *Шик А. Я.* Внутризонная фотопроводимость гетероструктур с квантовыми ямами // ФТП. 1986. Т. 20. С. 1598.
37. *Серженко Ф. Л., Шадрин В. Д.* Эффекты деполяризации в спектрах фотоионизации квантовых ям // Письма в ЖТФ. 1990. Т. 16, № 10. С. 34.
38. *Петров А. Г., Шик А. Я.* Эффекты деполяризации и фотоионизация квантовых ям // ФТП. 1997. Т. 31. С. 666.
39. *Ансельм А. И.* Введение в теорию полупроводников. М.: Наука, 1978.
40. *Грандштейн И. С., Рыжик И. М.* Таблицы интегралов, сумм, рядов и произведений. М.: Физматгиз, 1963.
41. *Кошляков Н. С., Глинер Э. Б., Смирнов М. М.* Основные дифференциальные уравнения математической физики. М.: Физматгиз, 1962.

42. *Bastard G.* Wave mechanics applied to semiconductor heterostructures. Les Ulis Cedex, 1988.
43. *Козырев С. В., Маслов А. Ю.* Влияние флуктуаций состава твердых растворов на подвижность двумерного электронного газа в полупроводниковых гетероструктурах // ФТП. 1988. Т. 22. С. 433.
44. *Spector J., Stoermer H. L., Baldwin K. W. et al.* Electron focusing in two-dimensional systems by means of an electrostatic lens // Allp. Phys. Lett. 1990. Vol. 56. P. 1290.
45. *Добрецов Л. Н., Гомоюнова М. В.* Эмиссионная электроника. М.: Наука, 1966.
46. *Ландау Л. Д., Лифшиц Е. М.* Квантовая механика, нерелятивистская теория. М.: Наука, 1989.
47. *Ландау Л. Д., Лифшиц Е. М.* Теория поля. М.: Наука, 1973.
48. Квантовый эффект Холла. Сборник статей. М.: Мир, 1989.
49. *Шкловский Б. И., Эфрос А. Л.* Электронные свойства легированных полупроводников. М.: Наука, 1979.
50. *Willett R., Einstein J. P., Stromer H. L. et al.* Observation of an even-denominator quantum number in the fractional quantum Hall effect // Phys. Rev. Lett. 1987. Vol. 59. P. 1776.
51. *Van Wees B. J., van Houten H., Beenakker C. W. J. et al.* Quantized conductance of point contacts in a two-dimensional electron gas // Phys. Rev. Lett. 1988. Vol. 60. P. 848.
52. *Kelly M. J.* Low-dimensional semiconductors. Oxford: University Press, 1995.
53. *Ford C. J. B., Simpson P. J., Pepper M. et al.* Coulomb blockade in small quantum dots // Nanostructured Materials. 1993. Vol. 3. P. 283.
54. *Кейси X., Паниш М.* Лазеры на гетероструктурах: М.: Мир, 1981. Т. 1, 2.
55. Ledentsov *N. N.* Quantum dot lasers: the birth and future trends // ФТП. 1999. Т. 33. С. 1039.
56. *Зайцев С. В., Гордеев Н. Ю., Копчатов В. И. и др.* Лазеры на квантовых точках: основные компоненты пороговой плотности тока // ФТП. 1997. Т. 31. С. 1106.
57. *Шейнкман М. К., Шик А. Я.* Долговременные релаксации и остаточная проводимость в полупроводниках // ФТП. 1976. Т. 10. С. 209.
58. *Абакумов В. Н., Перель В. И., Яссиевич И. Н.* Захват носителей заряда на притягивающие центры в полупроводниках // ФТП. 1978. Т. 12. С. 3.
59. *Luryi S.* Hot-electron transistors // High-speed semiconductor devices / Ed. S. M. Sze. New York: Wiley, 1990. P. 399.
60. *Thornton T. J.* Mesoscopic devices — what are they? Quantum transport in ultrasmall devices. NATO ASI School, 17—30 July, Il Ciocco, Italy, 1994. P. 65.
61. *Likharev K. K.* Single-electronics: correlated transfer of single electronics in ultrasmoll junctions, arrayes, and systems // Granular Nanoelectronics. NATO ASI Series 251. / Ed. D. K. Ferry, J. R. Barker, C. Jacaboni. 1990. P. 371.
62. *Smith C. G.* Low-dimensional quantum devices // Pep. Prog. Phys. 1996. Vol. 59. P. 235.
63. *Meirav U., Foxman E. B.* Single-electron phenomena in semiconductors // Semicond. Sci. Technol. 1995. Vol. 10. P. 255.
64. *Altmeyer S., Kühnel F., Spangenberg B., Kurz H.* A possible road to 77 K single-electron devices // Semicond. Sci. Technol. 1996. Vol. 11. P. 1502.

65. *Averin D. V., Korotkov A. N., Likharev K. K.* Theory of single-electron charging of quantum wells and dots // Phys. Rev. 1991. Vol. 44B. P. 6199.
66. *Odintsov A. A.* Single electron transport in two-dimensional electron gas system with modulated barriers: A possible *dc* current standard // Appl. Phys. Letts. 1991. Vol. 58. P. 2695.
67. *Kouwenhoven L. P., Johnson A. T., van der Vaart N. C. et al.* Quantized current in quantum dot turnstile using oscilling tunnel barriers // Phys. Rev. Lett. 1991. Vol. 67. P. 1626.
68. *Nakazato K., Blaikie R. J., Cleaver J. R. A., Ahmed H.* Single-electron memory // Electronics Lett. 1993. Vol. 29. P. 384.
69. *Lent C. S., Tougaw P. D., Porod W., Bernstein G. H.* Quantum cellular automata // Nanotechnology. 1993. Vol. 4. P. 49.
70. *Lent C. S., Tougaw P. D.* A device architecture for computing with quantum dots // Proc. IEEE. 1997. Vol. 85. P. 541.
71. *Lent C. S., Tougaw P. D.* Lines of interacting quantum-dot cells: A binary wire // J. Appl. Phys. 1993. Vol. 74. P. 6227.
72. *Tougaw P. D., Lent C. S.* Logical devices implemented using quantum cellular automata // J. Appl. Phys. 1994. Vol. 75. P. 1818.
73. *Snider G. L., Orlov A. O., Amlani I. et al.* Experimental demonstration of quantum-dot cellular automata // Semicond. Sci. Technol. 1998. Vol. 13A. P. 130.
74. *Korotkov A. N.* Wireless single-electron logic biased by alternating electric field // J. Appl. Phys. 1995. Vol. 78. P. 2520.

ОГЛАВЛЕНИЕ

Учебное издание

Александр Яковлевич Шик, Людмила Германовна Бакуева, Сергей Федорович Мусихин, Сергей Александрович Рыков

ФИЗИКА НИЗКОРАЗМЕРНЫХ СИСТЕМ

Редактор издательства *А. Л. Иванова*
Художник *Л. А. Яценко*
Технический редактор *Н. К. Румянцева*
Корректоры *Ю. Б. Григорьева* и *О. Ю. Гуршева*
Компьютерная верстка *И. Ю. Иняохиной*

Лицензия ИД № 02980 от 06 октября 2000 г. Сдано в набор 25.07.2001
Подписано к печати 26.09.2001. Формат 60 × 90 1/16. Бумага офсетная.
Гарнитура таймс. Печать офсетная. Усл. печ. л. 10. Уч.-изд. л. 9.2.
Тираж 1000 экз. Тип. зак. № 4250. С 204

Санкт-Петербургская издательская фирма «Наука» РАН
199034, Санкт-Петербург, Менделеевская лин., 1
main@nauka.nw.ru

Санкт-Петербургская типография «Наука» РАН
199034, Санкт-Петербург, 9 лин., 12